University of Alberta

Productivity Studies Using Advanced ANN Models

By

Ming Lu ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**

in

Construction Engineering and Management

The Department of Civil and Environmental Engineering

University of Alberta

Edmonton, Alberta, Canada

Spring 2001

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-60322-9

Canada

University of Alberta

Library Release Form

Name of Author: Ming Lu

Title of Thesis: Productivity Studies Using Advanced ANN Models

Degree: Doctor of Philosophy

Year this Degree Granted: 2001

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Ming Lu
#306, 9904, 90AV
Edmonton, Alberta, Canada
T6E 2T3

Date: 01/11/2000

University of Alberta

Faculty of Graduates Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Productivity Studies Using Advanced ANN Models** submitted by **Ming Lu** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Construction Engineering and Management**.

Dr. Simaan AbouRizk

Dr. Dave Sego

Dr. Samuel Ariaratnam

Dr. Ming Zuo

Dulcy Abraham

Dr. Dulcy Abraham

Soleymani

Dr. Hamid Soleymani

Date: Nov. 1. 2000

ABSTRACT

Estimating labor productivity is one of the most difficult aspects of preparing an estimate, or a control budget based on the estimate for labor-intensive activities in construction. The primary objective of research is developing artificial neural network or ANN based estimating tools to offer estimators valuable information about labor productivity in bidding new jobs.

In conjunction with a major Canadian industrial contractor, the thesis research presents case studies on the theoretical basis and practical considerations for measuring and analyzing labor productivity in industrial construction. Two important activities of process piping were investigated: pipe installation in the field and spool fabrication in the fabrication shop. Emerging computer modeling techniques such as data warehouses and ANN were researched from an academic perspective and implemented in industry to meet the challenges in productivity studies. The thesis research has addressed: (1) how to quantify labor productivity in industrial construction from a contractor's point of view; (2) how to measure actual labor productivity in industrial construction based upon on-site control practices; and (3) how to utilize ANN to analyze the variability of actual labor production rates and the sensitivity of identified influencing factors.

Using actual data, the proposed ANN models were proven to be effective in both risk analysis and sensitivity analysis of construction labor productivity. The developed data warehouses and ANN-based decision-support tools have been

implemented or are in the process of implementation at the involved company. The final results of the research not only assist estimators to improve the accuracy of estimating labor production rates for studied activities in bidding new jobs, but also offer the management a precise and integrated view of corporate productivity information spanning across many business divisions. The experience and lessons learned from the successful, productive and mutually beneficial collaboration between academia and industry in the thesis research will potentially benefit other university-industry joint research projects in the future.

PREFACE

This thesis is organized in a paper format, consisting of five main chapters and five appendices. Every chapter is an independent paper and can be read separately. However, all the chapters are logically coherent and pertinent to the theme of thesis. Each appendix is a user manual for one computer program that was developed in house in the thesis research. Chapter 1 overviews the whole thesis by introducing background information, problem statements, research objectives, methodologies used, and contributions achieved. Chapter 2 discusses a case study of industrial construction labor productivity, which depicts the settings of the research. Chapter 3 presents a probabilistic neural network classification model along with its application in estimating the production rates of field pipe installation. Chapter 4 presents a sensitivity analysis method of back propagation neural networks along with its application in estimating the production rates of shop spool fabrication. Chapter 5 summarizes what has been done thus far and recommends what to do in the future research. Appendix A is for the PINN trainer program based on the model described in Chapter 3. Appendix B is for the FabMaster program, which is the data warehouse for the fabrication facilities. Appendix C is for Fab_OLAP, which is an on-line analytical processing program in companion with FabMaster. Appendix D is for the PipingMaster program, which is the data warehouse for the field construction systems. Appendix E is for the SensitiveNN program based on the model as described in Chapter 4.

ACKNOWLEDGEMENT

First and sincerely, I would like to thank my university advisor, Dr. S. M. AbouRizk, without whose visions, guidance and encouragement this academic achievement would not have become a reality.

Especially, I would like to thank my industry advisor, Mr. U. H. Hermann of PCL, whose professionalism and enthusiasm have set the pace and the standard for the whole work. I am extremely grateful to PCL Industrial Constructors, Inc. for sponsoring the research financially and allowing me to use its actual data for developing problems and validating solutions throughout the thesis.

Finally, I would like to thank my wife, Duojia, for her understanding, love and assistance in making this thesis from thoughts to finish. I dedicate this work to her and our arriving Huan Huan.

Table of Contents

CHAPTER 1 INTRODUCTION **1**
  BACKGROUNDS 1
    Industrial Construction *1*
    Productivity Studies *2*
    Productivity Models *3*
    Artificial Neural Networks *5*
  PROBLEM STATEMENTS 8
    Productivity Studies *8*
    ANN Models *12*
  RESEARCH OBJECTIVES 15
    Productivity Studies *16*
    Probabilistic Neural Network Modeling *16*
    Sensitivity Analysis of Neural Networks *16*
  METHODOLOGIES 17
    Reviewing Literature to Recognize Issues *17*
    Identifying Factors from Brainstorming by Domain Experts *17*
    Using Data Warehouse to Quantitative Data *18*
    Questionnaire Survey *19*
    Computer Programming *21*
  ACADEMIC CONTRIBUTIONS 21
  INDUSTRIAL CONTRIBUTIONS 22
  CONCLUSIONS 23
  REFRENCES 23

CHAPTER 2 A CASE STUDY OF INDUSTRIAL CONSTRUCTION LABOR PRODUCTIVITY **28**
  INTRODUCTION 28
  INDUSTRIAL CONSTRUCTION 30
  FIELD PIPE INSTALLATION 32
    Productivity Quantification *32*
    Productivity Measurement *34*
    Input Factors *36*
    Probabilistic neural network modeling *40*
  SHOP SPOOL FABRICATION 42
    Productivity Quantification *43*
    Productivity Measurement *46*
    Input Factors *47*
    Sensitivity Analysis of Influencing Factors *51*
  CONCLUSIONS 57
  REFRENCES 58

CHAPTER 3 ESTIMATING LABOR PRODUCTIVITY USING PROBABILITY INFERENCE NEURAL NETWORK **60**
  INTRODUCTION 60
    Problem Domain *61*
    Review of NN Applications *64*
  PROBABILITY INFERENCE NEURAL NETWORK (PINN) MODEL 64
    Introduction of the PINN Model *64*
    Overview of the PINN Topology and Process *67*
    Data Pre-Processing *70*
    Output Zone Setup *74*
    Processing Elements (PE) at Kohonen Layer *74*
    NN Learning Process *75*

In-Zone Competition Strategy at Kohonen Layer *81*
PDF Approximator at Bayesian Layer *82*
Outputs at Output layer *84*
Sample of Recall Processing *84*
IMPLEMENTATION OF THE PINN MODEL 86
Validation of the PINN Model on Testing Data *87*
Sensitivity Analysis of the PINN Model *89*
CONCLUSIONS 93
REFERENCES 93
APPENDIX: *Desino's Method to Calculate "Conscience" Value* *95*

CHAPTER 4 SENSITIVITY ANALYSIS OF NEURAL NETWORKS IN SPOOL FABRICATION PRODUCTIVITY STUDIES **97**
INTRODUCTION 97
LITERATURE REVIEW 100
BPNN ALGORITHM AND INPUT SENSITIVITY 104
BPNN Algorithm *104*
Input Sensitivity Based on Normalized Data *107*
Input Sensitivity based on Original Data *110*
BPNN VS. REGRESSION ANALYSIS 114
STATISTICAL ANALYSIS OF INPUT SENSITIVITY 121
INDUSTRIAL APPLICATION 125
Spool Fabrication Basics *125*
Factor Identification and Data Collection *126*
BPNN Training and Sensitivity Analysis *129*
Model Testing and Validation *131*
CONCLUSIONS 133
REFRENCES 133

CHAPTER 5 CONCLUSION AND RECOMMENDATION **136**
SUMMARY OF RESEARCH 136
Productivity Studies and Data Collection *137*
Probabilistic Neural Network Modeling *137*
Sensitivity Analysis of Back Propagation Neural Networks *138*
Conclusion *140*
FUTURE RECOMMENDATIONS 140
Quantification of Textual / Descriptive data *140*
Optimization of NN Structure *142*
Sensitivity Analysis of PINN Model *143*
FINAL REMARKS 145

APPENDIX A USER'S MANUAL FOR PINN TRAINER **146**

APPENDIX B USERS' MANUAL FOR FabMASTER **159**

APPENDIX C USERS' MANUAL FOR FAB_OLAP **180**

APPENDIX D USERS' MANUAL FOR PipingMASTER **184**

APPENDIX E USERS' MANUAL FOR SensitiveNN **194**

List of Tables

Table 2-1: Sample of pipe installation unit labor rates 33

Table 2-2: Input factors to pipe installation productivity 39

Table 2-3: Sample of degree-of-difficulty factors for converting welds into units 44

Table 2-4: Explanatory factors to spool fabrication productivity 50

Table 3-1: Input Factors and Data Type of PINN Model 72

Table 3-2: Input Data Sample of PINN Model 73

Table 3-3: Scaled Input Vector and Initial Weight Vectors 78

Table 3-4: Updating Weight Vectors in First Learning Stage 79

Table 3-5: Updating Weight Vectors in Second Leaning Stage 81

Table 3-6: Trained PINN Ready to Recall for A Given Input Vector 85

Table 3-7: Recall Calculations at Bayesian Layer 86

Table 4-1: Data Set for Testing BPNN and Regression Analysis 117

Table 4-2: Partial Derivative (Slope) $(\frac{\partial N_1}{\partial N_P})$ at Four Input Points 119

Table 4-3: Statistics of Partial Derivative (Slope) Values: $(\frac{\partial N_1}{\partial S_P})$ 119

Table 4-4: Input Factors of Spool Fabrication Labor Productivity 128

Table 5-1: PINN vs. BP NN 144

Table B-1: Size Range Codes 159

Table B-2: Material Type Codes 160

Table B-3: Item Codes for Spool Level Data Compilation 160

Table B-4: Sample of FabMaster Outputs 170

Table D-1: Sample of Quantity Calculation Summary Table in PipingMaster 190

List of Figures

Figure 1-1: Sample Questionnaire for Finding Facts about Spool Fabrication 20

Figure 2-1: Output interface of PINN recall program 41

Figure 2-2: Sensitivity Analysis of Spool Fabrication BPNN Model.................... 54

Figure 2-3: Testing Sensitivity of BPNN to Material Type.................... 57

Figure 3-1: Topology of PINN Model.................... 66

Figure 3-2: Operations at Bayesian Layer in Recall.................... 80

Figure 3-3: Comparison of PINN and Back Propagation NN 88

Figure 3-4: PINN Output for the Base Case Scenario 89

Figure 3- 5: PINN Output for Scenario 1 91

Figure 3-6: PINN Output for Scenario 2 92

Figure 4-1: Structure of Back-Propagation NN Model 103

Figure 4-2: Illustration for Node and Layer Representations.................... 108

Figure 4-3: Distributions for Input Sensitivity.................... 120

Figure 4-4: Sensitivity Analysis of Spool Fabrication BPNN Model.................... 123

Figure 4-5: Testing Sensitivity of BPNN to Material Type.................... 132

Figure A-1: Select an identifier key of one previous trial 147

Figure A-2: User selects data table.................... 149

Figure A-3: Flag status of records.................... 150

Figure A-4: Setup structure and learning parameters for PINN 152

Figure A-5: Specify training iterations and train-test PINN 153

Figure A-6: Check training results 154

Figure A-7: Detected noise in training data.................... 155

Figure A-8: Global report for a train-test trial 156

Figure A-9: PINN Trainer on-line help.................... 157

Figure B-1: Program Flow Chart of FabMaster 165

Figure B-2: Main User Interface of FabMaster 128

Figure C-1: Select one ratio 181

Figure C-2: Trial on "number of pipe pieces per foot" 182

Figure C-3: View details of data 183

Figure D-1: Structures of Raw Data Tables for A Project 185

Figure D-2: Main user interface of FabMaster 186

Figure D-3: Handling: X-Reference Information Integrity Check 188

Figure D-4: Welding: X-Reference Information Integrity Check 189

Figure D-5: Productivity Analysis Page for Pipe Handling 191

Figure D-6: Sample of Pipe Handling Questionnaire 192

Figure D-7: Program Flow Chart of PipingMaster 193

Figure E-1: Splash Screen of SensitiveNN program 195

Figure E-2: Program Switchboard 197

Figure E-3: Open FFBPNN.mdb First 197

Figure E-4: Select data source table 198

Figure E-5: Examine details of data and edit record status 199

Figure E-6: Program main interface of SensitiveNN 201

Figure E-7: Check learning results when NN training terminates 202

Figure E-8: Check s Input Sensitivity for each input-output pair 204

Chapter 1: Introduction

BACKGROUND

Industrial Construction

Barrie et al (1992) described industrial construction as:

> "Industrial construction covers a wide range of construction projects that are essential to our utilities and basic industries, such as petroleum refineries and petrochemical plants, synthetic fuel plants, fossil fuel and nuclear power plants, off shore oil/gas production facilities, cryogenic plants etc. Industrial construction generally features large amounts of highly complex process piping, mechanical, electrical, and instrumentation work; both design and construction require the highest level of engineering expertise from multiple disciplines."

In particular, the installation of process piping systems in industrial construction is selected for productivity studies because it accounts for the bulk of direct labor hours of an industrial contractor. Process piping is used to transport fluids between storage tanks and processing units. Installation of piping systems generally consists of two processes: (1) spool fabrication in a commercial pipe shop; (2) pipe installation in the field. Although the two processes are inseparable and can be integrated to optimize the economics of a particular situation, they are treated independent of each other in the thesis because of the current estimating and control practices of the involved company. The productivity studies described in the thesis are conducted to support the management's decision-making in the context of the company's current management systems, as opposed to radically changing these systems.

Productivity Studies

In a construction task that is performed by hand labor, productivity is commonly expressed as the labor production rate (man-hours per installed unit), which measures a key dimension of performance and is a critical factor to estimating, scheduling and control of the project (Alfeld, 1988). Little information could be found in literature on the theoretical basis and practical considerations for measuring and analyzing labor productivity of industrial construction. In conjunction with a major industrial contractor (referred to as "the company" hereafter), productivity studies were conducted for two important activities in industrial construction: pipe installation in the field and spool fabrication in the fabrication shop.

In general, productivity studies encompass three tasks: (1) developing special methods and techniques to quantify labor productivity for estimating, and to measure actual labor productivity for on-site control, (2) identifying input factors that cause the variability in productivity, and (3) analyzing the relationships between input factors and productivity to enhance the accuracy of productivity estimating or improve the on-site performance directly. The focus of investigation is the average labor production rates (man-hours per unit) of these activities at the end of a project, rather than the daily labor production rates, because the primary objective of research is developing ANN-based estimating tools to offer estimators valuable information about labor productivity in bidding new jobs, rather than assessing and improving the crew performance in the field.

Productivity Models

Several established models for studying productivity can be found in the literature, including work study techniques, expectancy model, action-response model, regression model, expert systems, and artificial neural networks (ANN).

Work-study techniques were adopted in a number of productivity models, in which only a few factors related to work method were included (Thomas and Daily, 1983). Such work-study models cannot be used to model external and management factors. Thomas et al. (1990) and Thomas et al. (1991) discussed additional drawbacks of work-study techniques for construction productivity modeling.

The Expectancy model and action-response model are two alternative techniques proposed to explain variations in construction productivity. In the expectancy model, the effort that an individual is willing to exert accounts for the differences in job performance or productivity (Maloney and McFillen 1985). The action-response model graphically depicts the interaction of a number of factors that lead to the loss of productivity (Halligan et. al. 1994). Both models contribute to understanding the variations in productivity; however, neither can be used to quantify the influences of multiple factors on construction productivity (Sommez and Rowings, 1998).

Sanders and Thomas (1993) developed an additive linear regression model to study the effect of six project-related variables on masonry productivity based on data obtained from 11 projects. Eight binary variables were used in the model to represent the variations in productivity due to temperature and humidity. The effect of crew size was also taken into account in the model. The results of this regression model suggested higher productivity rates for crews with fewer members. Thomas and Sakarcan (1994) continued the research of Sanders and Thomas (1993) by developing the additive linear regression model for the purpose of forecasting labor productivity. They only included job condition variables that describe the work content and the physical components of the work. The focus of both studies was to determine the coefficient of condition variables, or the effect of a present condition on the activity productivity rate based on the results of historical study; such coefficients were derived independently of other inputs without accounting for combined effects. In addition, the determined coefficients are constants based upon the average values of historical data, and do not reflect the real situations in which the values of such coefficients may vary with the specific job conditions.

Expert systems is another technique applied to model labor productivity in two studies found in literature. Hendrickson et.al. (1987) developed an two-stage expert system named "MASON" to estimate activity durations for masonry construction. First, the maximum expected productivity was estimated. Next, this rate was adjusted for various characteristics of job or site. The maximum productivity estimates and the following adjustments were based on the knowledge obtained from interviews with a professional mason and a supporting laborer. Christian and Hachy (1995) developed an expert system to estimate the production rates for concrete pouring. The expert system relied on the knowledge extracted from experts and data collected from seven construction sites. The user simply queried the expert system for an estimate through a question-and-answer routine. In both expert systems, productivity was estimated through previously defined decision rules obtained from domain experts. Because the nature of formulating rules is subjective, the resultant rules may be inconsistent. Another disadvantage of analyzing productivity based on expert systems is expert systems do not perform functional input-output mapping, i.e. quantitative evaluation of the impact of job conditions on productivity.

In the following subsection background information about ANN models will be introduced and the technique of modeling productivity using ANN will be discussed.

Artificial Neural Networks

Artificial Neural Networks (ANN) research involves multiple disciplines including biology, artificial intelligence, computer science, and mathematics and evolves with the developments in each related discipline. Kohonen (1995) defined ANN as " massively parallel interconnected network of simple (usually adaptive) elements and their hierarchical organizations, intended to interact with the objects of the real world in the same way as the biological nervous systems do." Simply put, an ANN model is an analytical model that simulates the cognitive learning process of the human brain, and is automatically constructed from learning examples or data by trial and error without heuristic design or other human intervention.

ANN deals effectively with ill-structured problems, in which the algorithms required to solve them cannot be given in a precise and explicit fashion, or the data for a particular problem are either not complete or cannot be specified precisely (Widman et. al., 1989). ANN has been found to be capable of performing parallel computations on different tasks, such as pattern recognition, linear optimization, speech recognition, and prediction (Mukherjee and Deshpande 1995). In short, the special learning algorithms of ANN are capable of performing high dimensional, non-linear input-output mapping and extracting hidden patterns and predictive information from observing the learning examples.

In recent years, ANN has been researched and applied as a convenient decision-support tool in a variety of application areas in civil engineering, including modular construction decision making (Murtaza and Fisher, 1993), structural analysis (Flood and Katim, 1994), estimating construction productivity (Portas and AbouRizk, 1997), mode choice analysis of freight transport market (Sayed and Razavi, 1999), construction markup estimating (Li et al 1999), measuring organizational effectiveness (Sinha and McKim, 2000), and predicting settlement during tunneling (Shi, 2000).

In our research, ANN was selected as the main methodology and utilized to analyze the variability of actual labor production rates and the sensitivity of identified influencing factors due to two reasons.

First, construction labor productivity is influenced by a variety of factors. Model fitting based on construction labor productivity data requires quantification of the effects of factors on labor productivity and quantification of the interactions among the factors. The task of finding a mapping function from the independent variables to the dependent variable is analogous to that performed by some of the neural network models such as back propagation (Sonmez and Rowings, 1998). In statistics, regression analysis is the most common method to explore this relationship; in particular, the objectives and operations of nonlinear regression analysis are comparable to back propagation neural networks. However, regression models require the user to define a priori the parametric expression for the model (linear, quadratic, etc.). In the case of modeling productivity, the user is mainly concerned with what the productivity will be for any given set of work conditions, and may not necessarily be interested in the parametric expression of the model, for instance, a highly complex nonlinear functional equation. On the other hand, ANN is capable of nonlinear mapping for most complicated problems such as modeling productivity; the modeler does not need to exert much effort to decide on the class of relationships in a precise and explicit fashion.

Secondly, one of the attractive properties of ANN is their capacity for tolerating moderate amounts of noise in the data. In many real applications, the quantity and quality of the available data for modeling labor productivity may not support the fitting of a regression model. In such cases, ANN may be applied to generalize the knowledge from incomplete or noisy data and provide good solutions the problem.

Moselhi et al. (1991) pointed out the possible use of ANN for construction labor productivity modeling. Portas and AbouRizk (1997) developed an ANN model to estimate construction productivity for concrete formwork task. The majority of data used in the study was collected by questionnaires on a project basis. The prediction of the ANN model was compared with that of senior estimators for a single project. Sonmez and Rowings (1998) developed ANN models for quantitative evaluation of the impact of multiple factors on productivity in concrete pouring, formwork, and concrete finishing tasks, using data compiled from eight building projects. Their study also compared regression models including the pure linear regression model, the regression models with interaction and nonlinear terms with ANN models, and concluded, " the use of neural networks helped the overall modeling process. Neural networks have shown potential for quantitative evaluation of the effects of multiple factors on productivity, especially when interactions and nonlinear relations were present."

PROBLEM STATEMENTS

The problems to be solved in the thesis research were identified through investigating the current estimating and control practices of the involved company and reviewing the established ANN models and applications as found in the literature. Placed into two different perspectives, i.e. productivity studies and ANN models, the defined problems can be stated as follows:

Productivity Studies

ESTIMATING FIELD PIPE INSTALLATION PRODUCTIVITY

Estimating labor production rates for field pipe installation commences with establishing base production rates for various work items. Base production rates reflect the contractor's present labor productivity level under normal work conditions that are most often encountered in the field. The installation location is one of the major considerations for an estimator to define a classification of work conditions. For example, the base production rates of pipe installation are valid for the corresponding base classification only, in which the installation location is above ground up to 12 ft high. An estimator determines a degree-of-difficulty factor (often referred to as "multiplier" in the company) for each non-base classification to adjust the base rates up or down in order to reflect the unfavorable or favorable work conditions for the job being estimated. This is a subjective decision process, requiring substantial experience and skill on the part of the estimator to determine realistic production rates for the work conditions to be encountered. Empirical degree-of-difficulty factors for each classification of work conditions based on the installation location serve as a guide or tool to assist in deciding on such factors and can be found in the company's business manual. For example, the degree-of-difficulty factor for underground pipe installation (4 to 10 ft deep) is about two times the factor for aboveground pipe installation (up to 12 ft high), while the factor for pipe installation inside building at over 10 ft of height is about two times the factor for underground pipe installation (4 to 10 ft deep).

Historical piping productivity data of 66 projects was collected from the company and compiled into numeric format for analysis. The following two observations with regard to the actual degree-of-difficulty factors can be made from the historical data of the company: (1) the degree-of-difficulty factor for one classification of installation location may reveal a widespread distribution instead of a constant value as in the company's business manual; and (2) different classifications of installation location may end up with very close values of the degree-of-difficulty factor, not as distinguished as in the company's business manual.

The above observations are not initially expected and the explanation can be attributed to the fact that more factors exist, other than the location of installation, which contribute to the variability in labor productivity. In practice, an estimator may adjust the value of degree-of-difficulty factor in the business manual on a job based on experience and specific job conditions, and subjected to the approval of senior management. Barrie et al (1992) found that construction labor productivity may fluctuate wildly due to numerous factors that affect it, and many are highly qualitative in nature, including the effect of location and regional variations, the learning curve, work schedule and work rules, environmental effects, crew experience and management factors. Identification of input factors in the study of field pipe installation productivity was mainly based on Knowles (1997). A total of 36 input factors are considered relevant and used to redefine the classification of pipe installation. Those factors include both global project-level information and specific activity-level information.

ESTIMATING SHOP SPOOL FABRICATION PRODUCTIVITY

To estimate a fabrication project, a special "unitization" scheme is applied to quantify the various work items uniformly into an abstract unit of measure called "fabrication unit" or "unit" by weighting them for their degree of difficulty. A degree-of-difficulty factor is empirically determined for each weld, taking into account pipe diameter, wall thickness of pipe, weld type (butt weld, socket weld, saddle and lateral welds) and the time required to lay out and perform the weld. Quantity of non-welding work items such as cutting, beveling, handling pipe and fittings, installing supports are also converted into "units" by applying corresponding degree-of-difficulty factors in the scheme.

Once the total "units" for a project are determined, the focus of productivity study in spool fabrication is on the production rate directly (man-hour/unit). Similar to deciding on the degree-of-difficulty factor for a classification of work conditions in field pipe installation, deciding on unit labor rate for spool fabrication requires the experience and judgment of the estimator. The environmental effects and management factors are not considered as significant factors, as in the field productivity studies, because of the controlled shop environment, consistent policy and management personnel during the period of investigation. A total of 19 input factors are identified as affecting labor productivity of shop spool fabrication based on consultation with experienced estimators and shop superintendents in the company.

CHALLENGE IN PRODUCTIVITY STUDIES

It is not straightforward to create a conventional analytical model so as to accommodate the impacts of numerous factors on the target risky variable – degree-of-difficulty factor or production rate. It takes years of site experience and estimating practice for an estimator to develop his/her own mental model. The decision process relies heavily on individual's experiences and the results are often inconsistent reflecting the experience and disposition of the estimator.

ANN has been proposed by many as an alternative to streamline the estimating process and reduce the subjective nature of the work.

ANN Models

UNCERTAINTY ANALYSIS OF NN OUTPUT

The classic Back Propagation NN predicts a single value without giving any backup information on the risks of taking this value as correct. Observing the actual values for the degree-of-difficulty factors of field pipe installation indicates that the target risky variable lies over a relatively wide range. The result from an informal end-user survey showed that estimators are more comfortable to accept a decision support model with the capability of analyzing the uncertainty of its output. Thus, a probabilistic NN modeling approach that can predict a distribution or probability density function over the output range is preferred and has been researched.

Portas & AbouRizk (1997) proposed a feed forward back propagation neural network model for estimating construction production rates of formwork. The network outputs a single point prediction along with a number of output zones, with equal likelihood of the production rate being in any one zone. The output zones are symmetric and divided evenly across the range of likely production rate values. During training, the output zone with the output that coincides with the actual production rate is rewarded with a primary score of 1.0, representing strong certainty. A certain degree of fuzziness is considered by rewarding the 2 adjacent output zones with secondary scores of 0.5, representing weak certainty. All the other output zones are assigned a score of 0. Once the NN is trained and inputs are entered, the NN will predict a point value as well as the likelihood of production rates being within the output zones. This model achieved limited success due to the fact that the adopted back-propagation NN model is long on non-linear regression, but short on classification.

Specht (1991) revisited Probabilistic Neural Network (PNN) and General Regression Neural Network (GRNN) algorithms with the objective of integrating statistics and neural training. GRNN/PNN is a memory-based feed forward neural network model, where the training is performed in one pass, thus requiring less training time. GRNN/PNN is able to identify a posterior distribution over the NN weight vectors and a point-value prediction is generated based on the predicted distribution. However, based on experimentations and observations, GRNN/PNN is not quite tolerant of noisy data (inaccurate or incomplete records) and imposes a demanding standard of data quality that is hard to achieve in reality. The memory demand and computing time for GRNN/PNN increase very rapidly when the dimension of input vector and the quantity of training samples increase.

Kohonen proposed two special NN models, namely Self-Organizing Map (SOM) in the late 1980s and Learning Vector Quantization (LVQ) in the middle 1990s. SOM performs unsupervised classification and clustering to represent high-dimensional, nonlinearly related data items in an illustrative, often two-dimensional display. LVQ combines unsupervised and supervised leaning and is recommended for statistical pattern recognition problems. In LVQ, "decision surfaces, relating to those of the Bayesian classifier, are defined by nearest-neighbor classification with respect to sets of codebook vectors assigned to each class and describing it" (Kohonen, 1995). It is noted that the predicted result of LVQ and SOM is deterministic, being classified into one of *n* predefined clusters or classes.

Knowles and AbouRizk (1997) presented a two-stage NN model in predicting pipe-installation labor productivity. The input factors are used to invoke a LVQ classification process, followed by a predictive one. With the classification, the model predicts whether the output is likely in a typical or non-typical range. The proper feed-forward back-propagation network is then executed. The drawback of this method is that a build-up of errors occurs when the classification fails. For instance, if the classification accuracy is 90% at the first stage of NN, and the prediction accuracy at the second stage of NN is 85%, the prediction accuracy of the whole NN is only 76.5% (90% times 85%).

SENSITIVITY ANALYSIS OF NN INPUT

In contrast with a rather wide distribution of the actual production rate in field pipe installation, the actual labor production rates for shop spool fabrication are bounded within a relatively narrow range. Thus, the NN modeling of labor productivity in the shop puts more emphasis on the sensitivity analysis of influencing factors based upon the classic back propagation NN model, as opposed to the uncertainty analysis of expected production rate.

However, learning algorithms such as BPNN do not attempt to infer causality, hence, classification or prediction is based on blind correlation of new examples with previously analyzed examples, without giving information on the effect of each input parameter or influencing variable upon the predicted output variable. In the reported NN applications, model validation has thus far relied upon measuring accuracy of the calibrated network to an independent testing data set that are hidden from the neural network in learning. The model's sensitivity to changes in its parameters is generally probed by testing the response of a mature network on various input scenarios. In short, a NN model functions like a "black box" package, giving no clue on how the answers or model outputs are obtained, or how the input parameters affect the output.

Widman et. al (1989) pointed out that the credibility of an AI program frequently depends on its ability to explain its conclusions. Lack of interpretability is a pitfall of the neural network models recognized by many and has inhibited NN from achieving its full potential in real-world applications. Dhar and Stein (1997) argued that because NN algorithms such as the back-propagation NN are non-linear, high dimensional functional equations featuring parallel distributed data processing, it is hard to explicitly interpret which parameters cause what behavior in the NN model. While mathematical and operational methods do exist for the analysis of neural networks, the methods are fairly involved, and are less than satisfying because of their theoretical assumptions. They stated that "unlike most statistical methods, it can be difficult to say, even in general, which variables are significant in what respect." (Dhar and Stein 1997)

RESEARCH OBJECTIVES

The ultimate goal of the thesis research is to find better neural network modeling approaches to predicting production rates and productivity indices. When applied in industrial construction estimating as decision-support tools, the developed ANN-based models for analyzing productivity should be acceptable and effective to offer estimators valuable information about labor productivity in bidding new jobs. To attain this ultimate goal, the following objectives are defined in regard to three aspects:

Productivity Studies

Investigate the current estimating and on-site control practices for industrial construction as applied to the involved company, in order to advance the theoretical basis and practical considerations for measuring and analyzing labor productivity in industrial construction.

Probabilistic Neural Network Modeling

Building upon the previous developments achieved by others, establish a more effective NN approach that suits the needs of estimating industrial construction projects, which requires the recreation of a new training and recall algorithm that combines the functionality of probabilistic classification and prediction in one integrated neural network.

Sensitivity Analysis of Neural Networks

Define the input sensitivity of a NN model in mathematical terms, and establish a method of interpreting the relevance and impact of NN input parameters on the predicted output variable so as to gain insight into the rationale by which NN reason and make decisions.

METHODOLOGIES

Main methodologies utilized to fulfill the above research objectives include the following.

Reviewing Literature to Recognize the Issues

A comprehensive literature review was conducted in regard to the established ANN models, productivity studies, ANN applications in the problem domain, optimization, statistics, and industrial construction. Literature covers a wide range of journals, books, and reports, which document the latest academic developments and industrial applications in the related areas. Literature review helps recognize the issues to be addressed in the thesis, namely, how to get data from industry in modeling labor productivity, how to analyze the uncertainty of the output from an ANN-based productivity model, and how analyze the sensitivity of the input for an ANN-based productivity model.

Identifying Factors from Brainstorming by Domain Experts

The senior management and domain experts at the involved company including superintendents, production engineers, construction engineers, drafting superintendents, quality control superintendents, and welding foremen were convened for a brainstorming exercise to identify the factors that influence productivity of the studied activities. It should be noted that those factors as identified to influence labor productivity holds only within a specific setting and over a specific period. The input factors may need adjustment by adding relevant ones and deleting irrelevant ones when the setting of application changes to a different contractor, or a different period, even if the construction process being studied remains the same.

Using Data Warehouses to Gather Quantitative Data

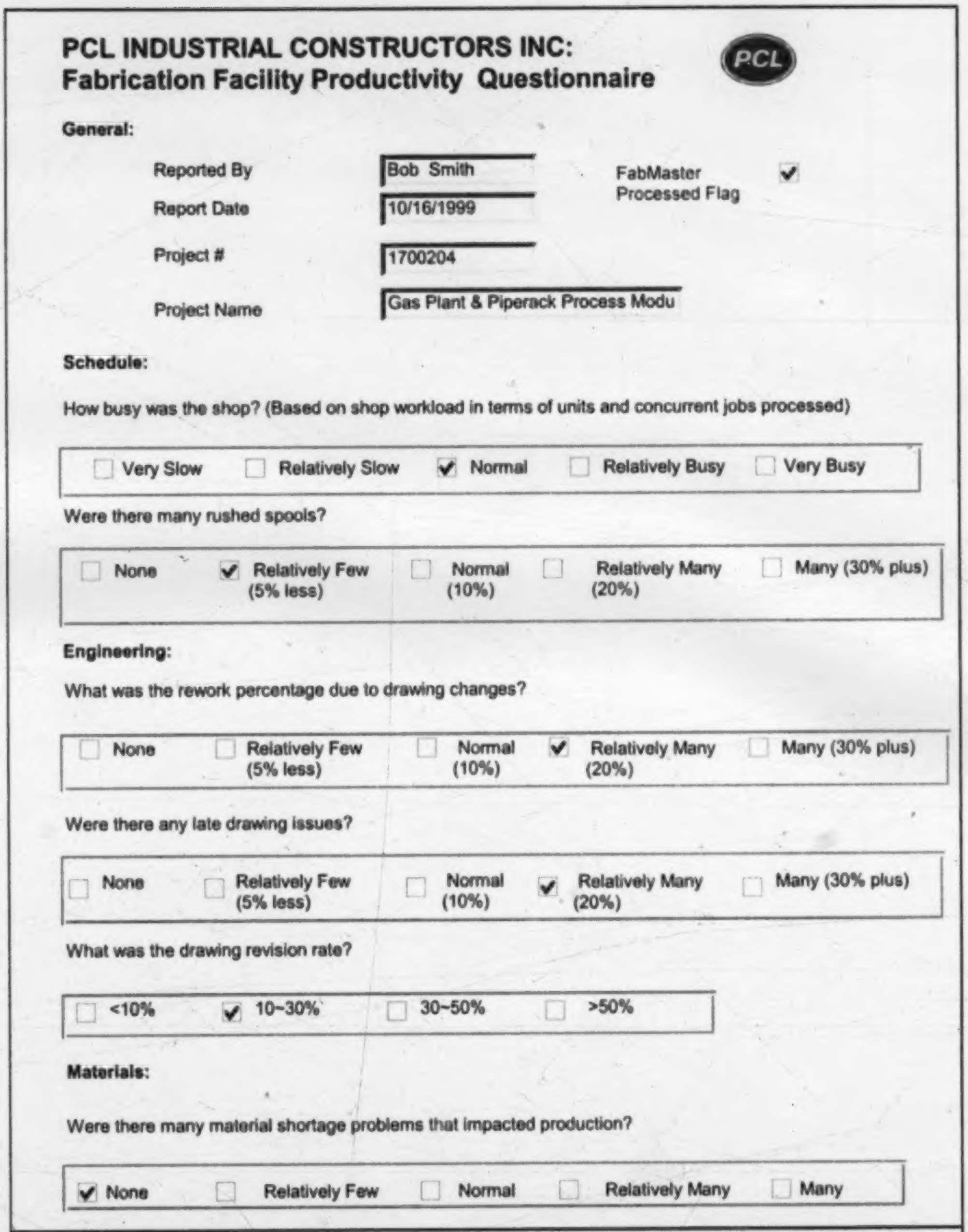
Identifying relevant factors and gathering high quality data for those factors are crucial to the success of modeling labor productivity using ANN. Following identification of factors, data needs to be collected.

The collaborative company (PCL Industrial) provided us with access to its business data for validation of ANN models and development of ANN-based decision-support tools. Although the company has invested resources in management information systems at various business divisions, those systems were developed and implemented separately. Productivity studies using ANN require vast amounts of data from different information management systems. A corporate data warehouse is "a process by which related data from many operational systems is merged to a single, integrated business information view that spans many business divisions" (Wang, 1997). With the support of the company's management, two data warehouses, namely, PipingMaster and FabMaster, were custom-developed for field pipe installation and shop spool fabrication respectively to integrate the corporate management systems of estimating, production resources planning, quality control, and labor cost control. Validating and processing of quantitative data were automated through computer programming within data warehouses. The developed data warehouses provide solid platform of integrated historical data from which to validate the ANN models and develop ANN-based tools for productivity analysis.

If data for some factors is not recorded in the existing management systems, questionnaire surveys were carefully designed and personnel at the company were interviewed to collect the needed data.

Questionnaire Survey

With the help of domain experts, questions and descriptive information of choices for ratings on a five-point scale were formulated into a questionnaire format with the objective of reducing ambiguities and confusions. It is worth mentioning that such questionnaire survey for modeling productivity using ANN is intended to find facts of the past projects only. Basically, in conducting the questionnaire survey, no personal judgment or opinion about the relationships between the facts and the results is involved. Questions of "What" type were asked about the factors affecting productivity only, and no questions of "Why" and "How" types were asked about the relationships between the factors and the productivity. In fact, ANN would sort out the relationships between the facts and the results on its own through an iterative learning process based on exploring sample data. Intelligence emerges when ANN finds the input-output patterns or relationships hidden in the data. This feature draws a distinct line between ANN approach and other intelligent modeling approach such as expert systems: ANN relies on facts and data, but requires less direct input from domain experts (Dhar and Stein, 1997). In short, modeling productivity using ANN is relatively an objective approach compared to expert systems. Figure 1-1 shows the questionnaire designed for finding additional facts about spool fabrication.



PCL INDUSTRIAL CONSTRUCTORS INC:
Fabrication Facility Productivity Questionnaire

General:

Reported By: Bob Smith
Report Date: 10/16/1999
Project #: 1700204
Project Name: Gas Plant & Piperack Process Modu

FabMaster Processed Flag ☑

Schedule:

How busy was the shop? (Based on shop workload in terms of units and concurrent jobs processed)

☐ Very Slow ☐ Relatively Slow ☑ Normal ☐ Relatively Busy ☐ Very Busy

Were there many rushed spools?

☐ None ☑ Relatively Few (5% less) ☐ Normal (10%) ☐ Relatively Many (20%) ☐ Many (30% plus)

Engineering:

What was the rework percentage due to drawing changes?

☐ None ☐ Relatively Few (5% less) ☐ Normal (10%) ☑ Relatively Many (20%) ☐ Many (30% plus)

Were there any late drawing issues?

☐ None ☐ Relatively Few (5% less) ☐ Normal (10%) ☑ Relatively Many (20%) ☐ Many (30% plus)

What was the drawing revision rate?

☐ <10% ☑ 10~30% ☐ 30~50% ☐ >50%

Materials:

Were there many material shortage problems that impacted production?

☑ None ☐ Relatively Few ☐ Normal ☐ Relatively Many ☐ Many

Figure 1-1: Sample Questionnaire for Finding Facts about Spool Fabrication

Following the formulation of a questionnaire, superintendents, project managers and estimators who were involved in the past projects were interviewed to compile facts and gather the needed information. The interview process was straightforward; the domain experts had no difficulty finding the records or recalling the facts on the projects that they managed.

Computer Programming

Microsoft Visual Basic, Visual Basic for Application, and Access were found to be flexible and powerful in handling large amounts of data and complex programming logic, hence, were selected as computer programming tools to develop both the data warehouses and ANN models in the thesis research. All the programs in this thesis research including data warehouses, ANN trainers, and ANN recall programs were developed in house without third-party software and have been utilized in the involved company.

ACADEMIC CONTRIBUTIONS

The solutions to the identified problems, which are provided through the thesis research, will contribute to the general knowledge of productivity studies and ANN modeling in regard to:

- Advancing the theoretical basis and practical considerations for measuring and analyzing labor productivity in industrial construction, which has been documented in a paper entitled "A case study of industrial construction labor productivity" and has been submitted for publication in the Journal of Construction Engineering and Management, ASCE;

- Devising a new neural network scheme to meet the requirements in modeling labor productivity of industrial construction, and is termed the Probability Inference Neural Network (PINN). PINN is a classification-prediction combined neural network model based on Kohonen's LVQ concept (Kohonen, 1995), but integrated with a probabilistic approach, which has been documented in a paper entitled "Estimating labor productivity using probability inference neural network" and is published in the October/2000 issue of the Journal of Computing in Civil Engineering, Vol 14(4), pp 241-248, ASCE;

- Establishing a simulation-based method of interpreting the relevance and impact of back propagation NN input parameters on the predicted output variable so as to gain insight into the rationale by which back propagation NN reason and make decisions, which has been documented in a paper entitled "Sensitivity analysis of neural networks in spool fabrication productivity studies" and has been submitted for publication in the Journal of Computing in Civil Engineering, ASCE.

INDUSTRIAL CONTRIBUTIONS

The developed data warehouses and ANN-based decision-support tools have been implemented or are in the process of implementation at the involved company. The final results of the research not only assist estimators in improving the accuracy of estimating labor production rates for studied activities in bidding new jobs, but also offer the management a precise and integrated view of corporate productivity information spanning across many business divisions. The experience and lessons learned from the successful, productive and mutually beneficial collaboration between academia and industry in the thesis research will potentially benefit other university-industry joint research projects in the future.

CONCLUSIONS

The problems addressed in the thesis research were identified through investigating the current estimating practices in industry and understanding the real concerns of industry professionals. Emerging computer-modeling techniques such as data warehouses and ANN were researched from an academic perspective in order to meet with the challenges in industry. The proposed novel ANN models and developed decision support tools were proven to be effective in both uncertainty analysis and sensitivity analysis of construction labor productivity; they were validated using real data from industry and successfully applied to assist estimators in deciding on labor production rates for new jobs.

REFERENCES

Alfeld, L. E. (1988). *Construction productivity – on –site measurement and management.*, McGraw-Hill , New York, NY.

Barrie, D. S., and Paulson, Jr., B. C. (1992). *Professional construction management, including C.M., design-construct, and general contracting. 3rd ed.*, McGraw-Hill, New York, NY.

Christian, J., and Hachey, D. (1995). "Effects of delay times on production rates in construction." *Journal of Construction Engineering and Management*, ASCE, 121(1), 20-26.

Gerwin, E. (1992). "Fabrication and installation of piping systems". *Piping Handbook*, 6th ed., Nayyar, M. L. eds., McGraw-Hill , New York, NY, 297-361.

Hendrickson, D., Martinelli, D., and Rehak, D. (1987). "Hierarchical rule-based activity duration estimation." *Journal of Construction Engineering and Management*, ASCE, 113(2), 288-301.

Knowles, P. (1997). *Predicting Labor Productivity Using Neural Networks.* Master of Sciences Thesis, University of Alberta, Edmonton, AB.

Lu, M., AbouRizk, S.M. and Hermann, U.H. (2000). "Estimating labor productivity using probability inference neural networks.", *J. of Computing in Civil Engineering*, Accepted October 1999, 14(4), 241-248.

Moselhi, O., Hegazy, T., and Fazio, P. (1991). "Neural networks as tools in construction." *Journal of Construction Engineering and Management*, ASCE, 117(4), 606-625.

Portas, J., and AbouRizk, S.M. (1997). Neural Network Model For Estimating Construction Productivity. *J. of Constr. Engrg. & Mgmt.*, ASCE, 123(4), 399-410.

Dhar, V. and Stein, R. (1997). *Intelligent Decision Support Systems: The Science of Knowledge Work.* Upper Saddle River, Prentice-Hall, Inc., New Jersey.

Flood, I., and Kartam, N. (1994). "Neural networks in civil engineering: systems and applications." *J. Constr. Engrg. And Mgmt.*, ASCE, 124(1), 18-23.

Halligan, D. W., Demsets, L. A., Brown, J. D., and Pace, C. B. (1994). "Action-response models and loss of productivity in construction." *Journal of Construction Engineering and Management*, ASCE, 120(1), 47-64.

Li, Y., Shen, L. Y., and Love, P.E.D. (1999). "ANN-based mark-up estimation system with self-explanatory capabilities." *Journal of Construction Engineering and Management*, ASCE, 125(3), 185-189.

Maloney, W. F., and McFillen, J. (1985). "Valence of and satisfaction with job outcomes." *Journal of Construction Engineering and Management*, ASCE, 111(1), 53-73.

Mukherjee, A., and Deshpande, J.M. (1995), "Modeling initial design process using artificial neural networks", *Journal Computing in Civil Engineering*, ASCE, 9(3), 194-200.

Murtaza, M.B., and Fisher, D.J. (1993), "Neuromex: Neural Network System for Modular Construction Decision Making", *Journal Computing in Civil Engineering*, ASCE, 8(2), 221-233.

Sander, S. R., and Thomas, H. R. (1993). "Masonry productivity forecasting model." *Journal of Construction Engineering and Management*, ASCE, 119(1), 163-179.

Sayed, T., and Razavi, A. (1999). "Comparison of neural and conventional approaches to mode choice analysis" *Journal of Computing in Civil Engineering*, ASCE, 14(1), 23-30.

Shi, J. (2000). "Reducing prediction error by transforming input data for neural networks", *Journal Computing in Civil Engineering*, ASCE, 14(2), 109-115.

Sinha, S. K. and McKim, R.A. (2000). "Artificial neural network for measuring organizational effectiveness.", *Journal Computing in Civil Engineering*, ASCE, 14(1), 9-14.

Sonmez, R. and Rowings, J. E. (1998). "Construction labor productivity modeling with neural networks." *Journal of Construction Engineering and Management*, ASCE, 124(6), 498-504.

Thomas, H. R. and Sakarcan, A.S. (1994). "Forecasting labor productivity using factor model." *Journal of Construction Engineering and Management*, ASCE, 120(1), 228-239.

Thomas, H. R. (1991). "Labor productivity and work sampling: the bottom line." *Journal of Construction Engineering and Management*, ASCE, 117(3), 423-444.

Thomas, H. R., Maloney, M.F., Horner, R.M., Smith, G.R., Handa, V.K., and Sanders, S.R. (1990). "Modeling construction labor productivity." *Journal of Construction Engineering and Management*, ASCE, 116(4), 705-725.

Thomas, H. R., and Daily, J. (1983). "Crew performance measurement via activity sampling." *Journal of Construction Engineering and Management*, ASCE, 109(3), 309-320.

Wang, C., B. (1997). *Techno Vision II*, McGraw-Hill, New York, N.Y.

Widman, L.E., and Loparo, K.A (1989). "Artificial intelligence, Simulation, and modeling: a critical survey", *Artificial intelligence, simulation, and modeling*. L.E.Widman, K.A. Loparo, and N.R. Nielsen, eds., John Wiley & Sons Ltd, New York, NY, 1-45.

Chapter 2: A Case Study of Industrial Construction Labor Productivity[1]

INTRODUCTION

In a construction task that is performed by hand labor, the labor production rate (man-hours per installed unit) measures a key dimension of performance and is a critical factor to estimating, scheduling and control of the project (Alfeld, 1988).

Thomas et al (1999) identified the complexity of the design and the project management as two factors that affect labor productivity and investigated the measurements of daily labor productivity in building construction including masonry construction, concrete formwork construction, and structural steel erection. They found that good project management and consistency in design complexity result in relatively constant daily labor production rates.

---

1 A version of this chapter has been submitted for publication. ASCE, Journal of Construction Engineering and Management.

A good correlation was also found between the final cumulative production rate (an index of the average labor performance over the entire project period) and the variance of daily production rates. For instance, their study of masonry construction observed "high variability in daily production rates on the poor performing projects due to disruptions in the work resulting from congestion, sequencing, lack of materials, etc" (Thomas et al, 1999).

Little information could be found in literature on the theoretical basis and practical considerations for measuring and analyzing labor productivity of industrial construction. In conjunction with a major industrial contractor (referred to as "the company" hereafter), we conducted a productivity case study for two important activities in industrial construction: pipe installation in the field and spool fabrication in the fabrication shop. The focus of investigation is the average labor production rates (man-hours per unit) of these activities at the end of a project, rather than the daily labor production rates as in Thomas 1999, because the primary objective of research is developing ANN-based estimating tools to offer estimators valuable information about labor productivity in bidding new jobs rather than assessing and improving the crew performance in the field. This paper intends to address: (1) how to quantify labor productivity in industrial construction from a contractor's point of view; (2) how to measure actual labor productivity in industrial construction based upon on-site control practices; and (3) how to utilize Artificial Neural Networks (ANN) to analyze the variability of actual labor production rates and the sensitivity of identified influencing factors.

The paper is organized as follows: important characteristics of industrial construction pertinent to productivity studies are first discussed in the "Industrial Construction" section. Next, the "Field Pipe Installation" section reviews the current estimating method, the present reporting and accounting systems for field pipe installation in the company, and summarizes the techniques for quantification and measurement of field pipe installation productivity. Further, the input factors that cause the variability in the productivity of field pipe installation are discussed, and a probabilistic neural network approach to modeling pipe installation productivity is overviewed. The subsequent section "Shop Spool Fabrication" shifts the focus of productivity studies to the fabrication facilities of the company, and summarizes the techniques for quantification and measurement of spool fabrication productivity. The input factors that affect the production rate of spool fabrication are identified, and an NN-based sensitivity analysis approach to modeling spool fabrication productivity is presented.

INDUSTRIAL CONSTRUCTION

Barrie et al (1992) described industrial construction as:

> "Industrial construction covers a wide range of construction projects that are essential to our utilities and basic industries, such as petroleum refineries and petrochemical plants, synthetic fuel plants, fossil fuel and nuclear power plants, off shore oil/gas production facilities, cryogenic plants etc. Industrial construction generally features large amounts of highly complex process piping,

mechanical, electrical, and instrumentation work; both design and construction require the highest level of engineering expertise from multiple disciplines."

In particular, the installation of process piping systems in industrial construction is selected for productivity studies because it accounts for the bulk of direct labor hours of an industrial contractor. Process piping is used to transport fluids between storage tanks and processing units. Installation of piping systems generally consists of two processes: (1) spool fabrication in a commercial pipe shop; (2) pipe installation in the field (Gerwin, 1996). Although the two processes are inseparable and can be integrated to optimize the economics of a particular situation, they are treated independent of each other in the paper because of the current estimating and control practices of the involved company. The productivity studies described in this paper are conducted to support the management's decision-making in the context of the company's current management systems, as opposed to radically changing these systems.

Parker et al (1984) distinguished industrial construction from heavy construction in that industrial construction does not require fleets of construction equipment and plant (such as scrapers, loaders, cranes and trucks etc) to handle basic materials (such as earth, rock, concrete and asphalt etc). They further pointed out that industrial construction "tends to be much more labor-intensive, though some of the largest hoisting and materials-handling equipment is also required" (Parker et al, 1984). An industrial contractor usually owns the equipment or rents it from a long-term supplier, thus, the technology and machinery adopted in construction can be considered invariable for a relatively long period of time. This feature lends the productivity studies of industrial construction to the unit-cost estimating method, which is commonly applicable to labor intensive work where "labor production rates must be independent of equipment use and vary among projects only because of differences in labor productivity" (Parker et. al., 1984). For instance, considering the bid item "Pour Concrete Floor" in building construction, to estimate the total cost in terms of labor hours, work quantities are taken off in square meters of floor, then multiplied by a labor production rate, i.e., the labor hours required on one square meter of floor. Analogously, for field pipe installation in industrial construction, the amount of work-in-place is usually counted in pipe footage; field productivity for pipe installation is measured in the form of unit rate, i.e. manhours per foot of installed pipe.

FIELD PIPE INSTALLATION

Pipe installation in the field involves "the physical placement of pipe / pipe subassemblies, valves, and other specialty items in their required final location relative to pumps, heat exchangers, turbines, boilers, and other processing units" (Gerwin, 1996)

Productivity Quantification

In practice, pipe is customarily identified by diameter of pipe (defined by nominal pipe size) along with wall thickness of pipe (defined by schedule number). Hence, the production rate of pipe installation can be determined by the diameter and wall thickness of pipe; the larger the diameter and the thicker the pipe, the more labor hours is required to install one foot of pipe. Table 2- 1 shows samples of the labor rates for handling and erecting straight run pipe (man-hours/ft) as found in the public source (Page and Nation, 1982).

Table 2-1: Sample of pipe installation unit labor rates (Source: Page and Nation, 1982)

Nominal Pipe Size (Diameter) (1)	Schedule Number (Wall Thickness) (2)	Base Labor Rate (MH/Ft) (3)
2	10	0.20
2	80	0.24
6	10	0.28
6	80	0.38
6	160	0.50
16	10	0.75
16	80	1.02
16	160	1.39
24	160	2.04

Estimating labor production rates for field pipe installation starts with establishing base production rates for various work items. Base production rates reflect the contractor's present labor productivity level under normal work conditions that are most often encountered in the field. The installation location is one of the major considerations for an estimator to define a classification of work conditions. For example, the base production rates of pipe installation are valid for the corresponding base classification only, in which the installation location is above ground up to 12 ft high. An estimator determines a degree-of-difficulty factor (often referred to as "multiplier" in the company) for each non-base classification to adjust the base rates up or down in order to reflect the unfavorable or favorable work conditions for the job being estimated. This is a subjective decision process, requiring substantial experience and skill on the part of the estimator to determine realistic production rates for the work conditions to be encountered. Empirical degree-of-difficulty factors for each classification of work conditions based on the installation location serve as a guide or tool to assist in deciding on such factors and can be found in the company's business manual. For example, the degree-of-difficulty factor for underground pipe installation (4 to 10 ft deep) is about two times the factor for aboveground pipe installation (up to 12 ft high), while the factor for pipe installation inside building at over 10 ft of height is about two times the factor for underground pipe installation (4 to 10 ft deep).

Productivity Measurement

In the context of pipe installation, keeping track of piping labor by individual fittings and pipe sections is economically impractical, if not impossible, to implement in the current field reporting system of the company. Alfeld (1988) argued that measuring labor productivity requires grouping similar accomplishments and separating dissimilar accomplishment on the job site. The cost control practice of the company for field pipe installation is described next.

At the end of a day, the foremen turn in time cards for their crews, charging the number of labor hours to a series of cost codes. The cost codes of field pipe installation for a particular project separate pipe fitters' hours by classifications of installation location. Thus, the total labor hours of pipe installation at various locations for one project can be readily retrieved from the field labor cost control system of the company. However, this is not the case for the amount of work accomplished. Large amounts of various work items along with variations in size and wall thickness of pipe cause the inclusion of details of work accomplished in the foreman's time cards to be impractical, such as the amount of work-in-place counted in footage by diameter and wall thickness of pipe, the screw joint or bolt-up connections and the valves and supports installations associated with the installed pipe. Fortunately, the detailed records about the amount of work accomplished can be obtained indirectly from the company's quality control system and estimating system. Thus, we can match the actual manhours with the work accomplished for one classification of installation location, in order to compute the actual degree-of-difficulty factor (ϕ) as given in Equation (1):

$$\phi = \frac{H}{\sum_{i=1}^{N} P_i \cdot Q_i} \qquad (1)$$

Where H is the actual labor hours charged to pipe installation in one classification of installation location,

N stands for the total number of work items contained in one classification of installation location,

P_i is the base labor rate for the i^{th} work item,

And Q_i is the actual quantity accomplished for the i^{th} work item.

Note that the estimating process described in the preceding subsection is actually to transform Equation (1) to compute the labor hours (H), simply by plugging the quantity take-off as read from construction drawings into the quantity term (Q_i) in (1). Hence, the task of estimating labor productivity boils down to determining the degree-of-difficulty factor (ϕ) accurately for a future project scenario. It is expected that a constant value of the degree-of-difficulty factor (or at least a narrow range) could be found for each classification from the company's historical records and should be close to the empirical value in the business manual.

Input Factors

Historical piping productivity data of 66 projects was collected from the company and compiled into numeric format for analysis. Because data is not well formatted or readily accessible, a data warehouse was developed first to integrate the contractor's estimating system, quality control system and labor-cost control system in order to ease the burden of data collection and ensure the high quality of collected data.

The following two observations with regard to the actual degree-of-difficulty factors can be made from the historical data of the company:

- The degree-of-difficulty factor for one classification of installation location may reveal a widespread distribution instead of a constant value as in the company's business manual;
- Different classifications of installation location may end up with very close values of the degree-of-difficulty factor, not as distinguished as in the company's business manual.

The above observations are not initially expected and the explanation can be attributed to the fact that more factors exist, other than the location of installation, which contribute to the variability in labor productivity. In practice, an estimator may adjust the value of degree-of-difficulty factor in the business manual on a job based on experience and job conditions, and subjected to the approval of senior management. Barrie et al (1992) found that construction labor productivity may fluctuate wildly due to numerous factors that affect it, and many are highly qualitative in nature, including the effect of location and regional variations, the learning curve, work schedule and work rules, environmental effects, crew experience and management factors. Portas and AbouRizk (1997) determined seven categories of activity factors and five categories of project performance factors to be relevant to the labor production rate of concrete formwork construction. Thomas et al (1999) identified the complexity of the design and the project management as two major categories that affect labor productivity of masonry construction. In regard to industrial construction, Knowles (1997) investigated a spectrum of explanatory factors to identify those that affect the productivity of pipe installation and pipe welding in the field.

Identification of input factors in this study was based on Knowles 1997, with the addition of three more factors, i.e. the contract type (lump sum or reimbursable), installation of miscellaneous fittings (flanges, specials, elbows etc.), and the on-site labor charging errors between the cost code of pipe installation and that of pipe welding (since pipe fitters and welders mostly work side by side). A total of 36 input factors are considered relevant and used to redefine the classification of pipe installation. Those factors include both global project-level information and specific activity-level information, as shown in Table 2- 2. Aside from location of installation, more activity-specific factors are considered such as material type of pipe, the installation of non-pipe components (valves, supports, and miscellaneous items), non-weld joints in installation (screw joints, bolt-ups), the quantities of installed pipe at different size ranges (small bore, medium bore, and large bore), the learning curve factor (total quantity of installed pipe in footage), the crew experience etc. Factors pertinent to project are also included, such as the effect of location and regional variations (project location, province/state), project type variations (project definition, contract type, and prefabrication percentage), work schedule and work rules (overtime and unionized), environmental effects (seasonal), management factors (superintendent and project manager) etc.

Table 2-2: Input factors to pipe installation productivity

ID	NN Input Factor	Remarks
1	Project Location	Urban, Rural, Camp Job
2	Administration	General Expense
3	Year of Construction	89~92, 93~94, 95~96, 97~99
4	Province/State	AB, SK
5	Contract Type	Reimbersable, Lump Sum
6	Client	an index derived from historical data
7	Engineering Firm	an index derived from historical data
8	Project Manager	an index derived from historical data
9	Superintendent	an index derived from historical data
10	Project Definition	Chemical, Cryogenic, Gas, Refining
11	Work Scope	Confined / Scattered
12	Project Type	Upgrade Shutdown, Grass Root etc.
13	Prefab/Field Work	Percentages for Prefabrication
14	Average Crew Size	<25, 25~50, 50~100, >100
15	Peak Crew Size	<25, 25~50, 50~100, 100~150, >150
16	Uninized	Yes, No
17	Equipment & Material	Equip.& Matl Cost/ Direct MH
18	Extra Work	Original Project Cost/Final Projeject Cost
19	Change Order	No. of Change Orders/Total Direct MH)
20	Drawing & Specs Quality	1 Poor 3 Average 5 Excellent
21	Location Classification	U/G on Site, Fab Shop, A/G on Site etc.
22	Total Quantity (Learning)	Total Quantity In DiaInFt
23	Installation Quantities	Qty for Size Ranges, <2", 2"~16", >16"
24	Material Type	Alloy,Carbon Steel, FRP/PVC,etc.
24	Method Of Installation	Percentages of Hand Rigging
26	Pipe Supports	No. of Pipe Supports/Foot of Pipe
27	Boltups	No. of Boltups/Foot of Pipe
28	Valves	No. of Valves/Foot of Pipe
29	Screwed Joints	No. of Screwed Joints/Foot of Pipe
30	Misc. Components	Install Misc.Components MH/Foot of Pipe
31	Welding Impact	Welding Multiplier (Miscoding on Site)
32	Season	Percentages of Winter & Summer Work
33	Crew Ability	1 Very Low, 3 Average 5 Very High
34	Site Working Conditions	1 Extreme Problems ~ 5 No Problem
35	Inspection, Safety & Quality	1 Extremely Detailed ~ 5 Highly Tolerant
36	Overall Degree of Difficulty	1 Very Low 3 Average 5 Very High

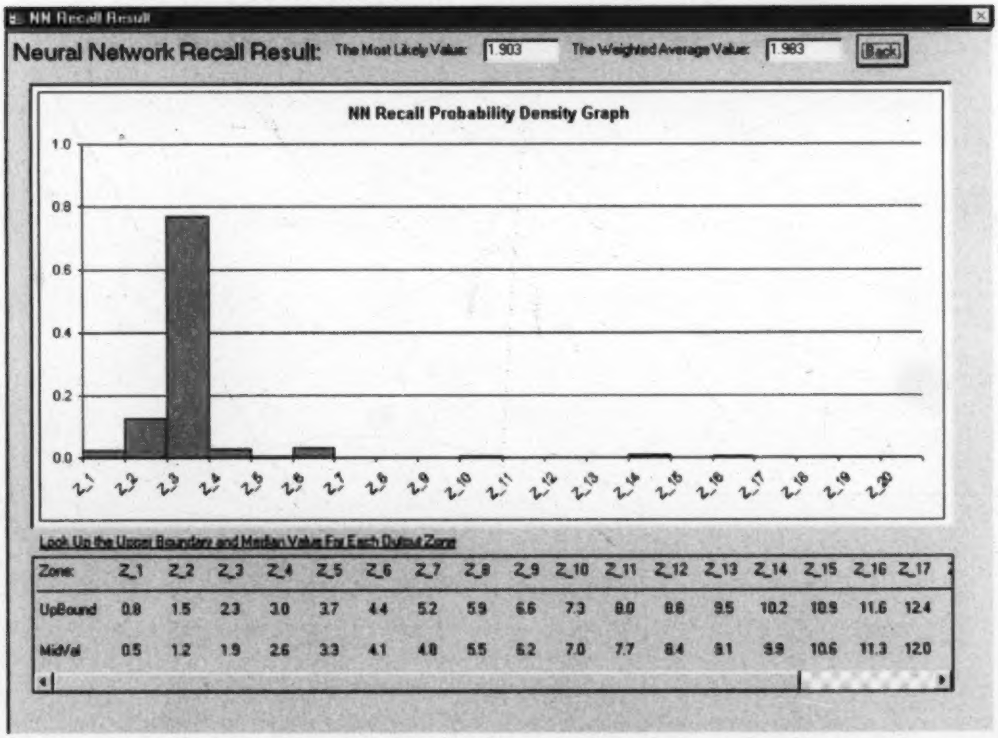
It should be mentioned that a questionnaire survey was carefully designed and conducted to collect some qualitative information that is not obtainable from the company's reporting and accounting systems. Such information was converted into numeric formats for the following NN analysis (See Lu et al, 2000 for details).

Probabilistic Neural Network Modeling

ANN has been proposed by many as an alternative to streamline the estimating process and reduce the subjective nature of the work. The classic Back Propagation NN predicts a single value without giving any backup information on the risks of taking this value as correct. Observing the actual values for the degree-of-difficulty factors of field pipe installation indicates that the target risky variable lies over a relatively wide range. The result from an informal end-user survey showed that estimators are more comfortable to accept a decision support model with the capability of analyzing the uncertainty of its output. Thus, a probabilistic NN modeling approach that can predict a distribution or probability density function over the output range is preferred and has been researched.

A new neural network scheme was devised to meet the requirements in modeling labor productivity of industrial construction, and is termed the Probability Inference Neural Network (PINN). PINN is a classification-prediction combined neural network model based on Kohonen's LVQ concept (Kohonen, 1995), but integrated with a probabilistic approach. Because the response of PINN is in the form of a probability density function (distribution) at the output range, an estimator will be able to decide on the degree-of-difficulty factor for a future scenario by combining the PINN's recommendation with personal judgment.

In the PINN model, the actual output range of the target risky variable is divided into a number of output zones or sub-ranges with an equal width. Output zones are actually some discrete clusters with continuous boundaries. For field pipe installation, the higher the value of the degree-of-difficulty factor, the higher the value of labor production rate, hence, the more difficult and more demanding the job is. Thus, each output zone gives an indication of the relative work difficulty and productivity level; for instance, output zone (0-0.7] stands for easier work and higher productivity level


NN Recall Result
Back
NN Recall Probability Density Graph
1.0
0.8
0.6
0.4
0.2
0.0
Z_1
Z_2
Z_3
Z_4
Z_5
Z_6
Z_7
Z_8
Z_9
Z_10
Z_11
Z_12
Z_13
Z_14
Z_15
Z_16
Z_17
Z_18
Z_19
Z_20

Neural Network Recall Result: The Most Likely Value: 1.903 The Weighted Average Value: 1.983

Look Up the Upper Boundary and Median Value For Each Output Zone

Zone:	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_10	Z_11	Z_12	Z_13	Z_14	Z_15	Z_16	Z_17
UpBound	0.8	1.5	2.3	3.0	3.7	4.4	5.2	5.9	6.6	7.3	8.0	8.8	9.5	10.2	10.9	11.6	12.4
MidVal	0.5	1.2	1.9	2.6	3.3	4.1	4.8	5.5	6.2	7.0	7.7	8.4	9.1	9.9	10.6	11.3	12.0

Figure 2-1: Output interface of PINN recall program

comparing with output zone (0.7-1.4]. The median of each sub-range can be used to represent the typical value for each output zone and to derive a predicted value in addition to the predicted distribution, such as mode and weighted average value.

Portable computer software was developed to implement the training and testing of the PINN model on real historical productivity data of field pipe installation at the company. The model was validated based on an independent data set reserved for testing. Sensitivity analysis of the model was performed by observing the PINN's output in response to controlled changes in inputs and comparing PINN's output against that of an experienced estimator. Following satisfactory testing and sensitivity analysis, a recall program based on the trained PINN model was implemented as a decision support tool for estimating the degree-of-difficulty factors of field pipe installation at the company. Figure 2- 1 shows the output interface of the recall program, indicating the predicted probability density function over the output range, and the likelihood of the degree-of-difficulty factor falling into each sub-range. Those who are interested in the topology and algorithm of the PINN model, and the effectiveness of applying PINN to estimate labor productivity in the context of field pipe installation may refer to Lu et al 2000.

SHOP SPOOL FABRICATION

Spool fabrication in a commercial pipe shop involves "the cutting, bending, tacking, and welding of individual pipe components to each other and their subsequent heat treatment and nondestructive examination to form a pipe subassembly or spool for installation" (Gervin, 1996). A pipe spool is a portion of piping system consisting of various piping components, such as flanges, elbows, reducers, tees, supports, and pipe. These components are prefabricated into distinct assemblies that are later assembled as part of an industrial plant or production skid/module. Such prefabrication is usually performed under controlled shop environment located away from the job site, which allows for better productivity and quality control, and hence cuts the field labor costs.

Major spool fabrication processes, such as cut, bevel, fit, weld, and handle sections of pipe and fittings, also tends to be labor-intensive. Productivity data is collected from the fabrication shop of the company for 63 projects completed from 1995 to 1999, during which period the technologies and machines for welding and cutting in the shop remain relatively stable. Like field pipe installation discussed previously, the productivity study of spool fabrication is suitable to the unit-cost estimating method.

Productivity Quantification

Alfeld (1988) pointed out the labor production rate in the shop could not be quantified with the same units as in the field – man-hours per foot of installed pipe, because the shop does not install the pipe but cuts, fits and welds spools; other units of measure such as spool counts and pipe sections do not satisfy the needs of quantifying the work accomplished in the shop either, because (1) each spool varies so much in components, size and configuration that a simple count of spools would be misleading; and (2) large-size pipe requires far more manhours to cut and weld than do the small-size pipe. Weld-inch was utilized as a unit of measure to quantify the accomplishment in a fabrication shop and Table 2- 3 shows samples of the degree-of-difficulty factors for converting various butt welds into weld inches as found in Alfeld, 1988.

Table 2-3: Sample of degree-of-difficulty factors for converting welds into units (Source: Alfeld, 1988)

Nominal Pipe Size (Diameter)	Circumference	Weld Type	Weighting Factors	Fab. Units
(1)	(2)	(3)	(4)	(5)
2.5	7.85	Butt	1.2	9.4
3.0	9.42	Butt	1.6	15.1
4.0	12.57	Butt	2.4	30.2
5.0	15.71	Butt	3.1	48.7

Similar to the concept in Alfeld 1988, in the fabrication shop of the company, a special "unitization" scheme is applied to quantify the various work items uniformly into an abstract unit of measure called "Fabrication Unit" or "Unit" by weighting them for their degree of difficulty. The "unitization" is a conversion based on a standard diameter inch along the circumference of a weld. A degree-of-difficulty factor is empirically determined for each weld, taking into account pipe diameter, wall thickness of pipe, weld type (butt weld, socket weld, saddle and lateral welds) and the time required to lay out and perform the weld. Quantity of non-welding work items such as cutting, beveling, handling pipe and fittings, installing supports are also converted into "Units" by applying corresponding degree-of-difficulty factors in the scheme.

A commercial fabrication shop usually handles several jobs simultaneously so that it is efficient for the crew to set up and do all the same size pipe from different jobs at the same time. In the fabrication shop of the company, it is difficult enough keeping track of the manhours charged to each individual job in the shop floor control systems. Charging labor hours to each individual pipe section or fitting is considered impractical and inefficient in light of the current control technologies and management systems in the fabrication shop.

The basic formula for spool fabrication estimating is shown in Equation (2):

$$H = P \cdot \sum_{i=1}^{N} (\phi_i \cdot Q_i) \qquad (2)$$

Where H is the total manhours charged to one job,

P is the production rate (MH/Unit) for the job,

N stands for the total number of work items (weld or non-weld) contained in the job,

Subscript i stands for the ith work item in the job,

ϕ_i is the degree-of-difficulty factor for the ith work item in the job,

Qi is the quantity for the ith work item in the job in its original unit of measure such as the weld counts for an weld work item, e.g. the weld count for “6 Nominal Pipe Size (Diameter), 40 Schedule Number (Wall Thickness), Butt-Weld Type” weld is 20.

Productivity Measurement

$\sum_{i=1}^{N}(\phi_i \cdot Q_i)$ in Equation (2) is actually the total quantity of fabrication work in Units for the job. The first step in estimating a spool fabrication job is a process called "unitization" for computing the total units of one job. The estimator reads the quantity takeoff from spool drawings and looks up the degree-of-difficulty factor for each work item. This task is straightforward but tedious because the amount of work items in a job is usually large; for example, several jobs the company completed contain over 1,000 spools, over 10,000 welds and over 10,000 pipe sections and fittings.

The difference in the degree-of-difficulty factor between field and shop should be noted:

- First, the degree-of-difficulty factor in the case of shop spool fabrication corresponds to each work item rather than a classification of grouped work items as in field pipe installation.

- Second, the degree-of-difficulty factors in the case of shop spool fabrication are held constant in the "unitization" scheme rather than variables as in field pipe installation.

Hence, the focus of productivity study in spool fabrication is on the production rate directly, i.e. the P term in (2) or man-hour/unit. Deciding on P requires the experience and judgment of the estimator. Similar to the productivity study of field pipe installation, a data warehouse was built up to integrate the reporting and accounting systems in the fabrication shop in order to obtain labor hours and quantity of fabrication work on each job. The data warehouse also contains a built-in computer program, developed to automate the tedious task of quantifying about 63 fabrication jobs into "units" in a precise and consistent way. Actual production rates over the period of investigation were observed for further analysis.

Input Factors

After consulting with experienced estimators and shop superintendents in the company, a number of quantitative and qualitative factors are considered relevant to the shop labor productivity, such as:

- The material components in fabrication, i.e. the percentage of non-carbon steel (stainless, aluminum, alloy steel etc.) units over the total units, because non-carbon steel spools require extra care and more time in storage, handling and welding comparing with carbon steel spools;

- The average length of pipe sections in a spool, indicated by in-line fittings (pieces) per foot of pipe in spool. In-line fittings, such as unions, couplings, swages, reducer etc are used to connect pipe sections in a straight line without turns or branches.

- The complexity of spool configuration, indicated by non in-line fittings (pieces) per foot of pipe in a spool, valves/supports/flanges (pieces) per foot of pipe in a spool;

- The stringency of quality control, indicated by the non-destructive test requirement, which is a percentage with respect to weld counts according to the client's specs.

- The quality of spool drawing indicated by the drawing revision rate.

- The shop workload, indicating shop's state of being busy or slow, and number of concurrent jobs handled at one time;

- The effect of double handling spools between weld stations, indicated by the percentage of multi-station roll weld inches over total roll well inches. A weld may be done on more than one station by different welders in the shop, depending on the welding process and the welder's qualification. It requires extra time to move spools between welding stations and lay out a weld at different stations.

- The effects of rushed spools due to client's priority, late drawing issues from the client, and material supply problems

- The amounts of night shift and overtime, and extra work in terms of labor hours;

- The experience and proficiency of crew, indicated by apprentice ratio, repair rate and reworked spools.

The environmental effects are not considered as significant factors, as in the field productivity studies, because of the controlled shop environment. A couple of management factors that were initially included were dropped out of analysis after examining the collected data, in which slight variations were observed due to the consistent management policy and management personnel during the 5-year period of investigation. It should also be mentioned that another factor describing the complexity of spool configuration was identified by domain experts, i.e. the number of pipe pieces per foot of pipe in spool. The sensitivity analysis results based on collected data reveal that the effect of the number of pipe pieces is very similar to that of the number of in-line fittings. Such strong correlation generalized by ANN model from the actual data is presented to domain experts and finds explanations from domain experts: pipe sections in a spool are mostly connected by in-line fittings such as unions, couplings, swages, reducer etc; both ratios, namely, in-line fittings (pieces) per foot and pipe pieces per foot, indicate the average length of pipe sections in a spool. To simplify the inputs of model, the ratio of pipe pieces per foot was dropped out of analysis, as agreed by domain experts. Eventually, nineteen input factors that affect labor productivity of shop spool fabrication are identified as listed in Table 2- 4.

Table 2-4: Explanatory factors to spool fabrication productivity

ID (1)	NN Input Factor (2)	Remarks (3)
1	In Line Fitting (pcs) per Foot of Pipe in Spool	A ratio indicating the average length of pipe sections in spool
2	Non In Line Fitting (pcs) per Foot of Pipe in Spool	A ratio indicating complexity of spool configuration
3	Valve (pcs) per Foot of Pipe in Spool	A ratio indicating complexity of spool configuration
4	Support (pcs) per Foot of Pipe in Spool	A ratio indicating complexity of spool configuration
5	Flange (pcs) per Foot of Pipe in Spool	A ratio indicating complexity of spool configuration
6	Multi-Station Roll Weld Inches / Total Roll Weld Inches	Multi-Station Roll Weld requires extra handling between weld stations
7	Repair Rate	An index of crew's proficiency
8	Radiography Test Requirement	An index of quality control stringency by specs.
9	Non CS Units / Total Units	Non CS component in fabrication requires extra care in storage, handling and welding
10	Shop Work Load	A 5-point rating based on shop workload in units and no. of concurrent jobs indicating how busy the shop was.
11	Drawing Revision Rate	A 5-point rating based on percent of revised spool drawings indicating drawing quality
12	Priority Rushed Spools	A 5-point rating based on percent of rushed spool due to client priority indicating shop work schedules.
13	Rework Spools	A 5-point rating based on percent of reworked spools due to drawing errors and quality defects
14	Material Shortage Problems	A 5-point rating on efficiency of material supply
15	Late Drawing Issues	A 5-point rating based on percent of late spool drawing issuance by client that impacts fabrication
16	Night Shift MHs / Total MHs	Night Shift affects labor productivity
17	Over Time MHs / Total MHs	Over Time affects labor productivity
18	Extra Work MHs / Total MHs	Extra Work affects labor productivity
19	Apprenticeship MHs / Total MHs	Welder qualification system affects labor productivity: Apprentice vs. Journeyman

Data for the identified factors is collected from the company's various management systems including labor cost tracking system, weld tracking system, payroll system, material tracking system. Because data is unavailable in current systems of the company for such factors as the material shortage problems, quantity of reworked spools, quantity of rushed spools due to priority, shop workload etc., a questionnaire survey was carefully designed and conducted with the support of the company management. The key personnel involved in the projects including shop superintendents, project managers and coordinators, QC staff, and welding foremen were interviewed to help recall some facts and gather the needed information.

Sensitivity Analysis of Influencing Factors

In contrast with a rather wide distribution of the actual production rate in field pipe installation, the actual labor production rates for shop spool fabrication are bounded within a relatively narrow range. Thus, the NN modeling of labor productivity in the shop puts more emphasis on the sensitivity analysis of influencing factors based upon the classic back propagation NN model, as opposed to the uncertainty analysis of expected production rate based on the PINN model.

Learning algorithms such as back-propagation NN do not give information on the effect of each input parameter or influencing variable upon the predicted output variable. The NN model's sensitivity to changes in its input factor is generally probed by testing the response of a mature network on various input scenarios. The relationships between an output variable and an input parameter were sorted out based on the NN algorithm so as to define the input sensitivity of a back-propagation NN model in exact mathematical terms in light of both normalized data and raw data (Lu et al, 2000). For a three-layer BPNN using Sigmoid transfer functions and linear normalization procedures, the input sensitivity with respect to the change of 10% input relevant ranges ($\frac{\partial N_n}{\partial S_p}^R$) is expressed as (3):

$$\frac{\partial N_n}{\partial S_p}^R = \frac{MAX_n - MIN_n}{10} \cdot \sum_{i=1}^{C} W_{pc_i} W_{c_i n} \cdot N_{c_i}(1 - N_{c_i}) \cdot N_n(1 - N_n) \qquad (3)$$

Where, subscript p stands for a node in the input layer of the network;

Subscript c stands for a node in the middle layer of the network; C stands for the total number of nodes in the middle layer;

Subscript n stands for a node in the output layer of the network.;

W_{ij} stands for the weight of connection between node i and node j;

S stands for the input signal to a node;

N stands for the output signal from a node;

MAX_n is the maximum value in the data set corresponding to output node n;

MIN_n is the minimum value in the data set corresponding to output node n.

From Equation (3), for a mature network, the sensitivity of an input parameter over an output variable is dependent on the current input values. A Monte Carlo simulation can be performed at the NN input space in order to observe the statistics of input sensitivity. In our research, statistical analysis of simulation results involves calculating 5 percentiles of the slope variable for each input parameter, i.e. the 10th, 25th, 50th, 75th, and 90th. The input sensitivity of all input parameters is summarized and presented in a tornado-like graph as illustrated in Figure 2- 2 for the piping fabrication labor productivity NN model. The horizontal axis represents the relative input sensitivity as determined by (3), i.e. output response (negative or positive) with a change of 10% relevant range in an input parameter. The vertical axis is the baseline corresponding to no output response or zero change in output. Five short vertical bars correspond to each input parameter, representing respectively the five percentiles from left to right in an ascending order and reflecting the central trend, the spread, and the shape of the observed slope data distribution from simulation. In short, statistical analysis of input sensitivity based on Monte Carlo simulation enables the modeler to understand the rationale of NN's reasoning and have pre-knowledge about the effectiveness of model implementation in a probabilistic fashion, as illustrated by the spool fabrication productivity model next.

A total number of 70 records were compiled and used to train a NN model with 19 input nodes at the input layer corresponding to 19 input parameters, 19 hidden nodes at the middle layer, and 1 output node at the output layer that is the unit labor hours. The number of hidden nodes can be determined based on trials; NN learning is found to be unsusceptible when to the number of hidden nodes is close to the number of input nodes. The learning rate is 0.4, the momentum is 0.1, and sigmoid transfer functions are used in hidden and output nodes. After satisfactory training (standard error of the output is 0.00143), the Monte Carlo based sensitivity analysis is performed on the matured network for 10000 simulation runs.

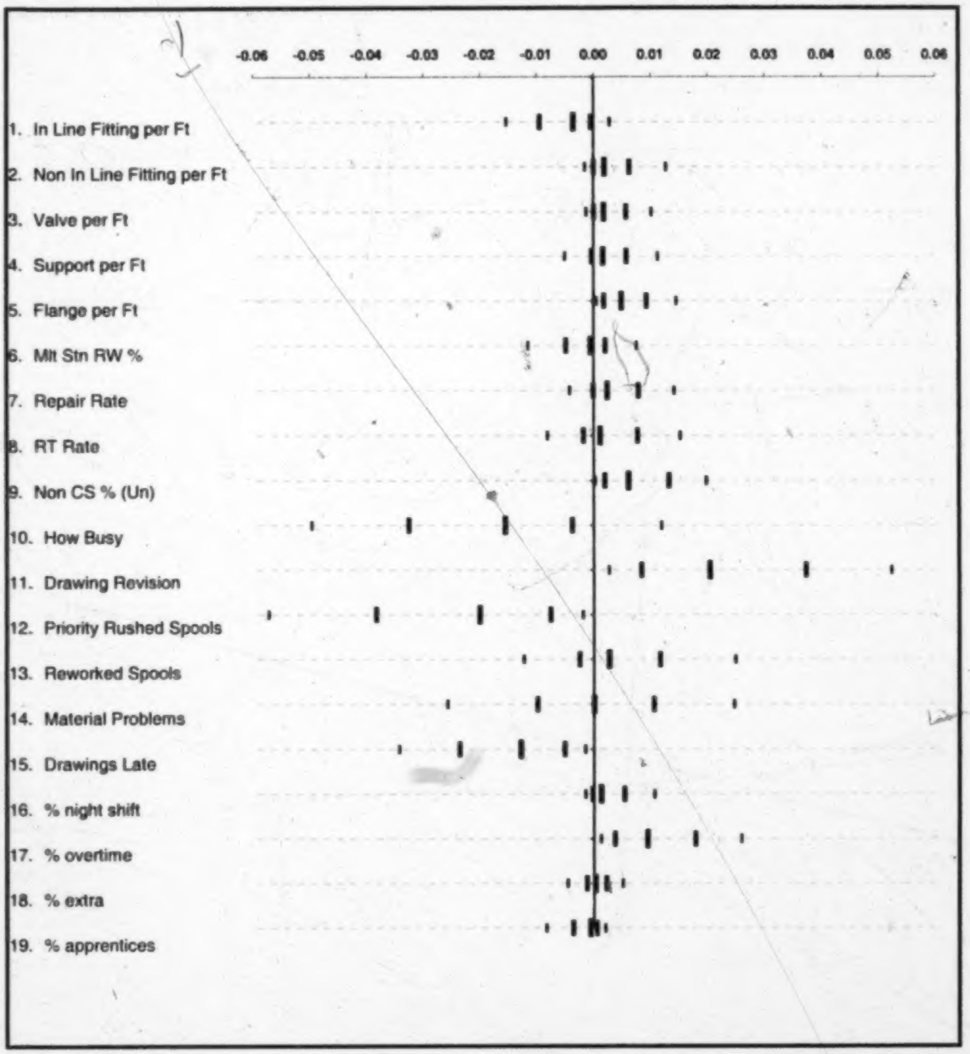


Figure 2-2: Sensitivity Analysis of Spool Fabrication BPNN Model

Several independent trials from NN training to the sensitivity analysis were conducted on the same data set. The best trial, in which the input sensitivity of most factors followed the same trends, as determined by experienced domain experts, is shown in Figure 2- 2. An examination of Figure 2- 2 reveals the relationships between the influencing factors and the fabrication productivity, which are generalized by NN through observing historical project data in the past 5 years. For example, factor 1 is about in line fitting pieces per foot of pipe in spool, which indicates the average length of pipe sections in spool. According to our domain experts, in line fittings, such as unions, couplings, swages, reducer etc are used to connect pipe sections in a straight line without turns or branches. Thus, the more in line fitting pieces in spools, the more small sections of pipe in spools, and the easier to handle the work. From Figure 2- 2, BPNN determines the chances to decrease labor hours per unit with the increase of this ratio are about 78% and agrees with the trend identified by domain experts. Factors 2 to 5 are four ratios indicating the complexity of spool configuration. By our domain experts, the higher such ratios, the more complex the spools' configuration, and the tougher to fabricate the spools. From Figure 2- 2, the dominant trends of the four ratios are all on the plus side, which matches the judgment of our domain experts. It is also observed from Figure 2- 2 that factor 18 (extra work percentage) is relatively tightly enveloped around the baseline, which indicates that extra work is not as dominant as other factors in contributing to the variance in unit labor rates. The explanation can be partly attributed to the fact that the amount of extra work impacts the efficiency of administration or management more directly than the productivity of crew on the shop floor. Other input factors can be interpreted and validated in a similar manner, and are not elaborated further due to space limit.

In particular, the effect of material type of spool fabrication on the labor productivity was tested based on the BPNN model, because material type (carbon steel, stainless steel, aluminum etc.) is a major consideration of an industrial estimator in adjusting unit labor hours of spool fabrication. The labor rate of non-carbon steel fabrication is empirically 1.5 times the rate of carbon steel in the company's business guideline. 24 records in the data set with 0% non-carbon steel component (100% carbon steel fabrication) were selected as testing records. Next, for each testing record, the input parameter of non-carbon steel component was changed from 0% to 100%, with other parameters intact. Those testing records were fed to the network and let NN recall the output, i.e. the unit labor rates for non-carbon steel fabrication. The output from NN was compared against the original output of each record, i.e. the unit labor rate for carbon steel fabrication. Based on the test results in Figure 2-3, NN increases the unit labor hours on 75% of the records; the amount of decrease for 5 records, i.e. No. 1, 2, 5, 6, 9, is relatively small comparing with the amount of increase for others. If the sample size is large enough, the percentage should come close to about 90%, as observed from Figure 2- 2 for factor 9. On average, the ratio of non-carbon steel labor rate over carbon steel labor rate is 1.4, which is close to 1.5 as in the guideline.

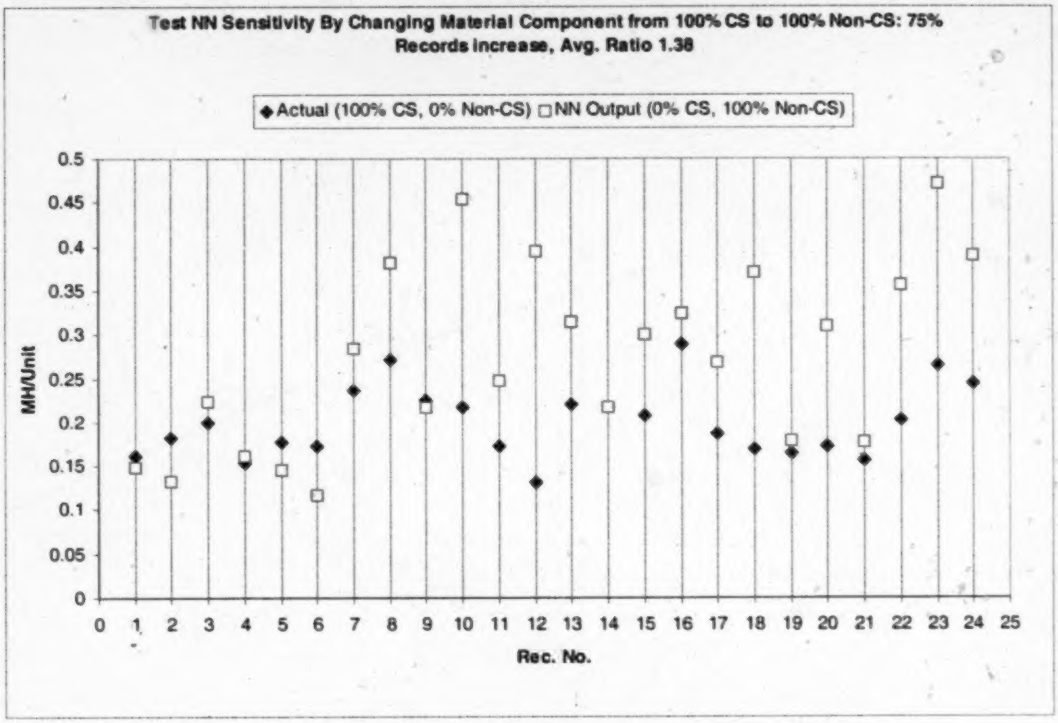


Figure 2-3: Testing Sensitivity of BPNN to Material Type

Note that the guideline gives an average number (1.5) in consideration of material type only, while NN is able to figure different numbers for different scenarios taking into account 19 relevant factors. In short, such a NN-based decision support tool will be more sophisticated and intelligent than the traditional business guideline to assist estimators in deciding on the labor production rate of spool fabrication.

CONCLUSIONS

Special methods are utilized in practice for the quantification and measurement of labor productivity in industrial construction. Estimating labor productivity is one of the most difficult aspects of preparing an estimate, or a control budget based on the estimate for labor-intensive activities in industrial construction. Artificial neural networks are capable of sorting out hidden patterns and extracting predictive information from complex data sets, and are proven to be effective in both uncertainty analysis and sensitivity analysis of construction labor productivity. The NN-based decision support tools are developed to assist estimators in deciding on labor production rates for new jobs; such tools can be more sophisticated and intelligent than traditional business manuals or guidelines.

REFERENCES

Alfeld, L. E. (1988). *Construction productivity – on –site measurement and management.*, McGraw-Hill , New York, NY.

Barrie, D. S., and Paulson, Jr., B. C. (1992). *Professional construction management, including C.M., design-construct, and general contracting. 3rd ed.,* McGraw-Hill , New York, NY.

Gerwin, E. (1992). "Fabrication and installation of piping systems". *Piping Handbook*, 6th ed., Nayyar, M. L. eds., McGraw-Hill , New York, NY, 297-361.

Knowles, P. (1997). *Predicting Labor Productivity Using Neural Networks.* Master of Sciences Thesis, University of Alberta, Edmonton, AB.

Lu, M., AbouRizk, S.M. and Hermann, U.H. (2000). "Estimating labor productivity using probability inference neural networks.", *J. of Computing in Civil Engineering*, ASCE, 14(4), 241-248.

Page, J.S. and Nation, J.G. (1982). *Estimator's piping man hour manual,* 3rd ed., Gulf Publishing Company Book Division, Houston,.

Parker, A.D., Barrie, D. S., and Snyder, R. M. (1984), *Planning and Estimating Heavy Construction*, McGraw-Hill, Inc., New York, NY.

Portas, J., and AbouRizk, S.M. (1997). Neural Network Model For Estimating Construction Productivity. *J. of Constr. Engrg. & Mgmt.*, ASCE, 123(4), 399-410.

Thomas, H. R., and Zavrski, I. (1999). "Construction baseline productivity: theory and practice." *J. Constr. Engrg. And Mgmt.*, ASCE, 125(5), 295-303.

Chapter 3: Estimating Labor Productivity Using Probability Inference Neural Network[1]

INTRODUCTION

Problem Domain

Estimating labor production rates (m-hr/unit) is both an art and a science. In general, the estimator develops the rate for a given project by starting with a "base rate" and modifying it to reflect the specific conditions he/she expects to encounter in the project being estimated. The base-rate is often determined statistically from past historical data, or from industry standards. In the context of industrial productivity estimating, the estimator accordingly adjusted the rate up or down by applying a difficulty multiplier to reflect overall favorable or unfavorable conditions. In determining the difficulty multiplier, consideration is only given to a couple of major factors that are thought to affect job productivity, such as installation location of pipe (inside a fabrication shop or on the job-site), and material type of welding (e.g. carbon steel or stainless steel).

---

[1] A version of this chapter has been published. ASCE, Journal of Computing in Civil Engineering, October/2000, Vol 14(4), pp 241-248.

The challenges of this approach include the fact that it is not straightforward to create a conventional mathematical model so as to accommodate the impacts of numerous factors on the target risky variable. The decision process relies heavily on individual's experiences and the results are often inconsistent reflecting the experience and disposition of the estimator.

Artificial neural networks have been proposed by many as an alternative for streamlining the process and reducing the subjective nature of the work. Most models, however, were based on point predictions of production rates with which estimators were uncomfortable. The point prediction by NN can be defined as a single value predicted by neural network models without any backup information on the risks of taking this value as correct. The new NN model presented in this paper arises out of the need for accurate prediction in the form of a distribution at the output range. The estimators will be able to make a decision for a future scenario based on the results recalled by the NN model and personal preferences and experiences.

In the following section, previous NN applications in the problem domain are first reviewed.

Review of NN Applications

Moselhi, Hegazy, and Fazio (1990) cite the prediction of a realistic productivity level for a certain trade as an aspect of construction that can be modeled with neural networks. Factors such as job size, building type, overtime work and management conditions are typically considered by an estimator and can easily be manipulated for use as neural network inputs.

Karshenas and Feng (1992) analyzed earth-moving equipment productivity with a neural network application. A modular neural network structure was used to make it possible to add specifications of new equipment with only a brief training session. Each module represents a distinct type of equipment that was trained with two inputs, four hidden nodes, and one output within a back propagation training algorithm.

Wales and AbouRizk (1993) used neural networks as a means of applying the effects of environmental site conditions to the labor production rate on an activity. Daily average temperature, precipitation, and cumulative precipitation over the previous seven days were identified as three key environmental site conditions and used as inputs into a feed forward back propagation neural network training algorithm. The output was a productivity factor such that a value larger than 1.0 indicates that environmental site conditions produce a greater than average productivity. On the other hand, a productivity factor of less than 1.0 indicates that the environmental site conditions result in below average productivity.

Chao and Skibniewski (1994) performed a case study in which a neural network was used to predict the productivity of an excavator. They identified two main factors that affect an excavator's productivity: job conditions and operation elements. Job conditions include the characteristics of the environment such as soil conditions, and specific characteristics of the excavator and excavation such as the vertical position of the cutting edge. Operational elements, in contrast, include characteristics not directly related to the excavating operation; for example, the effect of wait time for trucks and extra tasks other than excavating. Two neural networks were used for the purpose of this case study. The first was used to estimate the excavator cycle time. Four key factors were identified as having an influence: cycle time (including swing angle), horizontal reach, vertical position, and soil type (job conditions). The output of the first network was then incorporated into the second network, which examined the effect of the operational elements on the productivity.

Portas & AbouRizk (1997) proposed a feed forward back propagation neural network model for estimating construction production rates of formwork. The network outputs a single point prediction along with a number of output zones, with equal likelihood of the production rate being in any one zone. The output zones are symmetric and divided evenly across the range of likely production rate values. During training, the output zone whose output coincides with the actual production rate is rewarded with a primary score of 1.0, representing strong certainty. A certain degree of fuzziness is considered by rewarding the 2 adjacent output zones with secondary scores of 0.5, representing weak certainty. All the other output zones are assigned a score of 0. Once the NN is trained and inputs are entered and the NN will predict a point value as well as the likelihood of production rates being within the output zones. This model achieved limited success and its limitation was overcome by the work discussed next.

Knowles (1997) presented a two-stage NN model in predicting pipe-installation labor productivity. The input factors are used to invoke a LVQ classification process and then a predictive one. With the classification, the model predicts whether the output is likely in a typical or non-typical range. The proper feed-forward back-propagation network is then executed. The drawback of this method is that a build-up of errors occurs when the classification fails. For instance, if the classification accuracy is 90% at the first stage of NN, and the prediction accuracy at the second stage of NN is 85%, the prediction accuracy of the whole NN is only 76.5% (90% times 85%). This problem motivated the development of the model described in this paper by taking a different probabilistic approach, which is more direct and more meaningful in terms of giving a point prediction and quantifying its associated probability.

PROBABILITY INFERENCE NEURAL NETWORK (PINN) MODEL

Introduction of the PINN Model

Specht (1991) revisited Probabilistic Neural Network (PNN) and General Regression Neural Network (GRNN) algorithms with the objective of integrating statistics and neural training. GRNN/PNN is a memory-based feed forward neural network model, where the training is performed in one pass, thus requiring less training time. GRNN/PNN is able to identify a posterior distribution over the NN weight vectors and a point-value prediction is generated based on the predicted distribution. However, based on experimentations and observations, GRNN/PNN is not quite tolerant of noisy data (inaccurate or incomplete records) and imposes a demanding standard of data quality that is hard to achieve in reality. The memory demand and computing time for GRNN/PNN increase very rapidly when the dimension of input vector and the quantity of training samples increase.

The PINN model uses similar topology as the GRNN/PNN model but is a refinement of it. This is because PINN generalizes the underlying statistical patterns within training data and codes those patterns into a limited number of weight vectors through iterative learning. As a result, the number of weight vectors will not proportionally increase with the increase of dimensionality and quantity of training data.

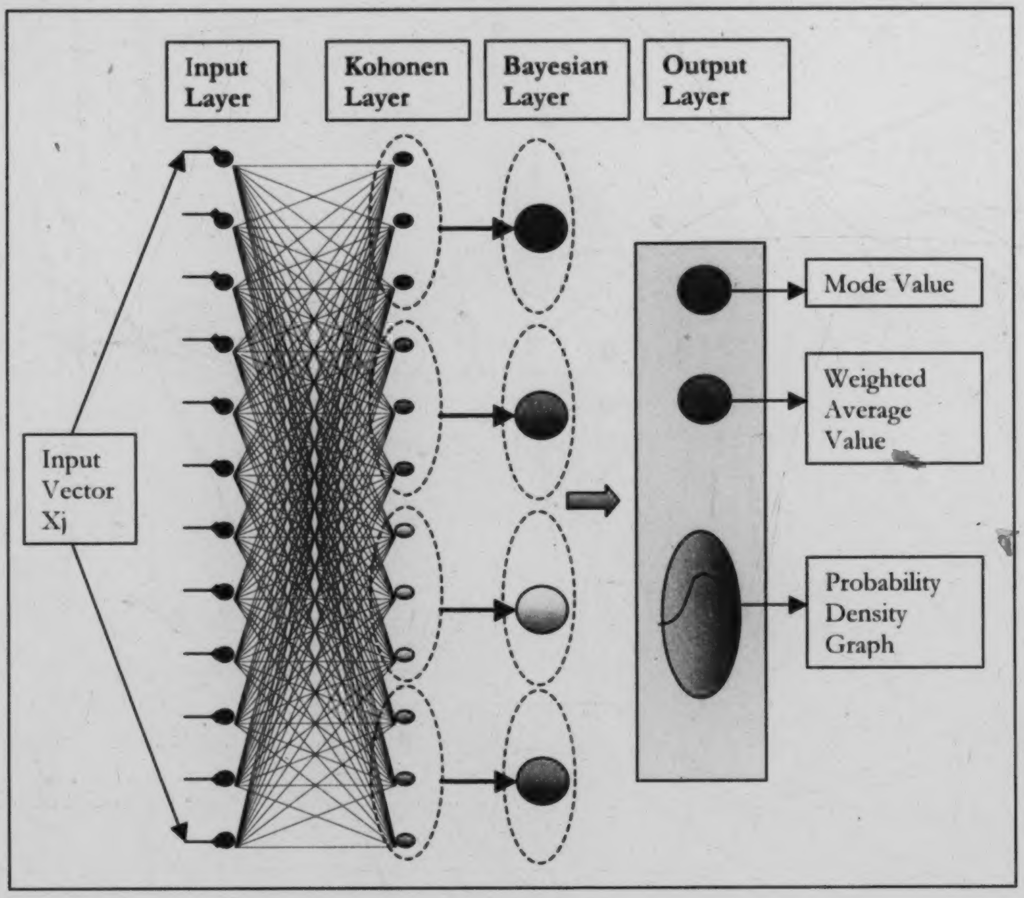

Input Layer
Kohonen Layer
Bayesian Layer
Output Layer
Input Vector Xj
Mode Value
Weighted Average Value
Probability Density Graph

Figure 3-1: Topology of PINN Model

The PINN model imbeds the output zone concepts described in Portas & AbouRizk (1997). In the application domain of industrial labor productivity estimating, the profile of actual historical productivity data reveals a spread range. The range of NN output value, i.e. the production rate multiplier, is evenly divided into a number of sub ranges, or output zones, which are actually some discrete clusters with continuous boundaries. The higher the multiplier value, the more difficult and more demanding the job is, and the lower the productivity for the job. Thus, each output zone gives an indication of the relative work difficulty and productivity level, for instance, output zone [0-0.7] stands for easier work and higher productivity level comparing with output zone [0.7-1.4]. The median of each sub range can be used to represent the typical value for each output zone and to derive a NN predicted value.

The PINN uses the same strategy as the model described in AbouRizk et. al. (1999) by incorporating Kohonen's LVQ in NN learning. The main difference is that the classification and prediction networks are combined in an integrated network, which required the development of a different training and recall algorithm. Murtaza and Fisher (1993) utilized Kohonen's unsupervised learning algorithm called self-organizing map or SOM for modular construction decision making. Kohonen's Learning Vector Quantization (LVQ) combines unsupervised and supervised leaning and is recommended for statistical pattern recognition problems (Kohonen, 1995). Three options for the LVQ-algorithms (LVQ1, LVQ2, and LVQ3) were proposed. Kohonen's research shows that each of the three LVQ variations yields similar accuracy in most statistical pattern recognition tasks, although a different philosophy underlies each algorithm. LVQ1 was utilized in the learning process of the PINN model.

Overview of the PINN Topology and Process

The topology of PINN model is given in Figure 3- 1. It is composed of four layers. The middle layers are a Kohonen classifier and a Bayesien layer. The outcome of the PINN model at the output layer is a probability density function or a distribution reflecting the likelihood of the target variable occurring in a given zone. The mode of the distribution and its mean can serve as point predictions.

The PINN process consists of four stages as follow:

(1) Preparation, which deals with

- Scaling data at the input layer, which will be discussed in the subsection titled "Data Pre-Processing";
- Setting up output zones at the Kohonen layer, which will be addressed in the subsection titled "Output Zone Setup"; and
- What are Processing Elements and how they function at the Kohonen layer, which will be discussed in the subsection titled "Processing Element at the Kohonen Layer".

(2) Learning, which takes place between the Input layer and the Kohonen layer using the LVQ algorithm. This does not involve the Bayesian layer or the Output layer. This will be discussed in the subsection "NN Learning Process".

Once learning is achieved, the input-output patterns are coded into the weight vectors of the processing elements at the Kohonen layer.

(3) Investigating whether the neural network has been successfully trained. This is accomplished through the following steps:

- Feed the input vectors of the training and testing records into the input layer of the PINN model.

- Project the input vector of one record onto the Kohonen layer by using the results of stage (2). The Euclidean distances between each processing element's weight vector and the scaled input vector are calculated.

- An in-zone competition occurs within every output zone at the Kohonen layer, which is detailed in the subsection titled "In-Zone Competition Strategy at Kohonen Layer". The processing element with the minimum Euclidean distance value wins.

- Project the winner PE at the Kohonen layer onto the Bayesian layer. The Bayesian layer holds a probability density function (PDF) approximator. The Euclidean Distance values of the winner PEs are the inputs to the PDF approximator. The following subsection of "PDF Approximator at Bayesian Layer" discusses the components and operations in more details.

- The output is mapped from the Bayesian layer and presented in the form of a probability density function at the output layer. Two point predictions are calculated in addition to the probability density function namely, the mode and the weighted average. The subsection "Outputs at the Output Layer" includes details about the NN outputs.

- Check the NN outputs against the actual outputs of the training and testing records. If the results are satisfactory, then the neural network is declared to have been trained; otherwise, repeat stage (1) using different parameters at each layer.

(4) Recall. Once the model calibration is done, the neural network can be used to recall the output value for any given input vector, which is similar to stage (3) using the final results determined in stage (2) and (3). A sample calculation is given in the subsection "Sample of Recall Process".

Data Pre-Processing

At the input layer of the PINN model (shown by Figure 3- 1), the number of input nodes corresponds to the dimension of the input vector. The dimension of the input vector depends on the number of input factors and the input data types. Three input data types are used to define NN input factors, i.e. "Raw", "Rank", and "Binary". "Raw" is used simply for quantitative input factors, like general expense ratios, winter construction percentages, or quantities of work. "Rank" is used to convert subjective factors, like crew ability ratings, into numeric format. And "Binary" is used to group textual factors into numeric formats, like material type and project definition. It should be noted an input factor of the "Raw" or "Rank" type corresponds to one input node at the input layer, while an input factor of the "Binary" type corresponds to a number of input nodes depending on the number of groups for the factor. For illustration, input factors and data types for the "Pipe Installation" neural network model are listed in Table 3-1. A sample record for the "Pipe Installation" NN training is also listed in Table 3-2 showing both the raw data and converted NN input data. The NN input data is normalized and scaled between 0 and 1 at the input layer. These scaled inputs will be passed forward for NN training. At the Kohonen layer all weight vectors are randomly initialized between 0 and 1.

Table 3-1: Input Factors and Data Type of PINN Model

NN Input Factor (1)	Data Type (2)	Options & Remarks (3)
Project Location	Binary	Urban, Rural, Camp Job
Administration	Raw	General Expense
Year of Construction	Binary	89~92, 93~94, 95~96, 97~99
Province/State	Binary	AB, SK
Contract Type	Binary	Reimbersable, Lump Sum
Client	Raw	an index derived from historical data
Engineering Firm	Raw	an index derived from historical data
Project Manager	Raw	an index derived from historical data
Superintendent	Raw	an index derived from historical data
Project Definition	Binary	Chemical, Cryogenic, Gas, Refining
Work Scope	Binary	Confined / Scattered
Project Type	Binary	Upgrade Shutdown, Grass Root etc.
Prefab/Field Work	Raw	Percentages for Prefabrication
Average Crew Size	Binary	<25, 25~50, 50~100, >100
Peak Crew Size	Binary	<25, 25~50, 50~100, 100~150, >150
Uninized	Binary	Yes, No
Equipment & Material	Raw	Equip.& Matl Cost/ Direct MH
Extra Work	Raw	Original Project Cost/Final Projeject Cost
Change Order	Raw	No. of Change Orders/Total Direct MH)
Drawing & Specs Quality	Rank	1 Poor 3 Average 5 Excellent
Location Classification	Binary	U/G on Site, Fab Shop, A/G on Site etc.
Total Quantity (Learning)	Raw	Total Quantity In DiaInFt
Installation Quantities	Raw	Qty for Size Ranges, <2", 2"~16", >16"
Material Type	Binary	Alloy,Carbon Steel, FRP/PVC,etc.
Method Of Installation	Raw	Percentages of Hand Rigging
Pipe Supports	Raw	No. of Pipe Supports/Foot of Pipe
Boltups	Raw	No. of Boltups/Foot of Pipe
Valves	Raw	No. of Valves/Foot of Pipe
Screwed Joints	Raw	No. of Screwed Joints/Foot of Pipe
Misc. Components	Raw	Install Misc.Components MH/Foot of Pipe
Welding Impact	Raw	Welding Multiplier (Miscoding on Site)
Season	Raw	Percentages of Winter & Summer Work
Crew Ability	Rank	1 Very Low, 3 Average 5 Very High
Site Working Conditions	Rank	1 Extreme Problems ~ 5 No Problem
Inspection, Safety & Quality	Rank	1 Extremely Detailed ~ 5 Highly Tolerant
Overall Degree of Difficulty	Rank	1 Very Low 3 Average 5 Very High

Table 3-2: Input Data Sample of PINN Model

NN Input Factor (1)	Raw Data (2)	NN Input Data (3)
Project Location	Rural	0 0 1
Administration Requirement	0.235	0.235
Year of Construction	94~95	0 1 0 0
Province/State	Alberta	1 0
Contract Type	Reimbersable	1 0
Client	Shell	1.037
Engineering Firm	Colt	0.682
Project Manager	John Doer	1.036
Superintendent	Bob Smith	0.982
Project Definition	Chemical	1 0 0 0 0
Work Scope	Confined to Specific Area	1 0
Project Type	Plant Upgrade No Shutdown	0 1 0 0
Prefab/Field Work	10%, 90%, 0%	10, 90, 0
Average Crew Size	25~50	0 1 0 0
Peak Crew Size	50~100	0 0 1 0 0
Uninized	Yes	1 0
Equipment & Material	9.4	9.4
Extra Work	0.661	0.661
Change Order	0.019	0.019
Drawing & Specs Quality	Excellent	5
Activity Location Classification	Inside <10ft High	0 0 0 1 0 0
Total Quantity (Learning)	6055	6055
Installation Quantities	210, 905, 4940	210, 905, 4940
Material Type	Carbon Steel	0 0 1 0 0
Method Of Installation	Hand Rigging%, Machine Rigging %	10, 90
Pipe Supports	0.45	0.45
Boltups	4.77	4.77
Valves	1.59	1.59
Screwed Joints	0	0
Misc. Components	3.18	3.18
Welding Impact	1.25	1.25
Season	Winter%, Summer%	10, 90
Crew Ability	Average	3
Site Working Conditions	Many Problems	2
Inspection, Safety & Quality	Detailed	2
Overall Degree of Difficulty	Low	2

Output Zone Setup

As discussed in the section "Introduction of the PINN model", the likely range of output values is evenly divided into a number of output zones. The output zone boundary setup is important for PINN learning and recall. Wide zones are generally not helpful to the decision-maker and hence should be avoided. Zones that are unacceptably tight may prevent PINN from learning. It requires some trials to obtain reasonable output zone boundaries and the following two aspects should be considered:

1. Precision requirement of the user, i.e. the zone width or sub-range that suffices for the user to make decisions.
2. Distribution of actual output data over the output zones. A uniform distribution of actual output data over all zones generally yields better results.

Processing Elements (PE) at Kohonen Layer

At the Kohonen layer, each output zone contains an equal number of processing elements. Each processing element is associated with a weight vector (also referred to as a codebook vector (Kohonen, 1995)).

Visually, a weight vector is a set of links that emanate from one processing element and end at each input node as illustrated in Figure 3- 1. Thus the dimension of a weight vector is equal to that of the input vector (the number of input nodes). An output zone at Kohonen layer can be visualized as a chip containing a number of pins (PEs).

During NN training the orientation of those pins is gradually fine-tuned to capture the underlying statistical patterns within the training data (Kohonen, 1995).

Our experience indicates that the number of processing elements assigned to each class should be close to the average frequency in the histogram of training data output values, i.e. the average number of training samples in one output zone.

NN Learning Process

Data of all the training records is scaled at the input layer. The scaled input data is fed into the model to calibrate the weight vectors of the Processing Elements (PE) at the Kohonen layer, using the LVQ algorithm suggested by Kohonen (1995).

The learning process involves a number of iterations, each of which is comprised of the following:

1. The Euclidean distances between the input vector of a training record and each PE's weight vectors are calculated. The PE that has the smallest Euclidean distance value is declared to be a global winner. If the global winner PE does not belong to the same output zone that the actual output value of this training record falls into, the weight vector of the global winner PE is penalized according to the following equation (1):

$$\mathbf{W}_{ij}' = \mathbf{W}_{ij} - RR \times (\mathbf{X}_j - \mathbf{W}_{ij}) \qquad (1)$$

Where:

RR stands for “Repulsion Rate”, which is a learning rate to penalize the global winner PE_i.

$\mathbf{X}_j^r$ is the input vector of the training record, and

$\mathbf{W}_{ij}$ is the current weight vector of the global winning PE_i.

$\mathbf{W}_{ij}'$ is the updated weight vector of the global winning PE_i.

The repulsion rate is initially set between 0 and 1, and is reduced gradually until it approaches 0 at the end of learning.

2. Following the global competition, an in-zone competition among processing elements occurs only at the output zone into which the actual output value of the training record falls. Prior to the in-zone competition, a “conscience” value is added to each PE's Euclidean distance and adjusted over the learning iterations, so as to effectively prevent one PE within a specific output zone from winning all the time and activate as many PEs as possible in the learning process. The formulas to calculate the conscience Euclidean distance can be found throughout the pertinent literature. The interested readers can refer to Appendix II for details.

The method we adopted is as follows:

The “conscience” Euclidean distance between each PE's weight vector and the input vector is calculated. The PE with the shortest “conscience” Euclidean distance value is declared to be an in-zone winner. Only the in-zone winner PE is rewarded using equation (2):

$$\mathbf{W}_{ij}' = \mathbf{W}_{ij} + AR \times \left(\mathbf{X}_j - \mathbf{W}_{ij}\right) \qquad (2)$$

Where:

AR stands for "Attraction Rate", which is a learning rate to reward the in-zone winner PE_i.

$\mathbf{X}_j$ is the input vector of a training record, and

$\mathbf{W}_{ij}$ is the current weight vector of the in-zone winner PE_i.

$\mathbf{W}_{ij}'$ is the updated weight vector of the in-zone winning PE_i.

Like the repulsion rate, the attraction rate is initially set between 0 and 1, and is reduced gradually until it approaches 0 at the end of learning iterations.

A sample calculation of one learning iteration is presented next to illustrate the learning process.

As shown in Figure 3- 1, the dimension of input vector is 12, and the output range ([0-4]) is divided into 4 output zones, i.e. [0-1], [1-2], [2-3], [3-4]. Each output zone contains 3 processing elements. Note that this sample is a simple model and serves for illustration. Problems encountered in a real situation, which are suitable for the PINN model to solve, are mostly high dimensional; the number of input nodes may exceed 100.

The input vector of a training record is scaled between 0 and 1, and the weight vector of a processing element at the Kohonen layer is randomly initialized between 0 and 1. Table 3-3 shows the input vector $\mathbf{X}_1$ and the weight vectors of the 3 processing elements in zone 1.

Table 3-3: Scaled Input Vector and Initial Weight Vectors

Input Vector $\mathbf{X}_1$ (1)	PE_1 $\mathbf{W}_{11}$ (2)	PE_2 $\mathbf{W}_{21}$ (3)	PE_3 $\mathbf{W}_{31}$ (4)
0.3	0.7232	0.0844	0.0819
0.2	0.6772	0.9129	0.2596
0.7	0.8780	0.4422	0.9704
0.8	0.6307	0.8774	0.8196
0.4	0.7523	0.7261	0.3969
0.5	0.8575	0.0439	0.3109
0.5	0.2088	0.4567	0.3809
0.9	0.6310	0.4542	0.0177
0.8	0.6065	0.2413	0.9379
0.1	0.8804	0.9188	0.8970
0.4	0.8232	0.7465	0.1775
0.6	0.0485	0.1158	0.7543

The Euclidean distance (ED) is calculated between the input vector $\mathbf{X}_1$ and each weight vector $\mathbf{W}_{i1}$ as $ED_{11} = 1.4163$, $ED_{21} = 1.5746$, and $ED_{31} = 1.2963$. Suppose that PE3 in zone 1 is the global winner PE among all the processing elements, which gives a minimum ED value of 1.2963. If the actual output of this training record does not fall into zone 1, i.e. outside the sub-range [0-1], then the weight vector of PE_3 ($\mathbf{W}_{31}$) is updated by equation (1) as shown in Table 3-4. Notice that the Repulsion Rate is set to be 0.8 at the start of learning in the sample calculation. Kohonen (1995) recommends smaller initial value such as 0.06 for the Repulsion Rate and Attraction Rate for obtaining better results.

Table 3-4: Updating Weight Vectors in First Learning Stage

Input Vector X_1 (1)	PE_1 W_{11} (2)	PE_2 W_{21} (3)	PE_3 W_{31}' (4)
0.3	0.7232	0.0844	-0.0927
0.2	0.6772	0.9129	0.3073
0.7	0.8780	0.4422	1.1867
0.8	0.6307	0.8774	0.8353
0.4	0.7523	0.7261	0.3943
0.5	0.8575	0.0439	0.1597
0.5	0.2088	0.4567	0.2857
0.9	0.6310	0.4542	-0.6881
0.8	0.6065	0.2413	1.0482
0.1	0.8804	0.9188	1.5346
0.4	0.8232	0.7465	-0.0004
0.6	0.0485	0.1158	0.8777

If the actual output of this training record does fall into zone 1, i.e. within the sub-range [0-1], then no weight vector is updated in the global competition. The learning process steps into the second phase.

In the first training iteration, the conscience value for every processing element in zone 1 is determined to be equal to 0 (see Appendix I). So the "conscience" Euclidean Distance value is equal to the original Euclidean Distance value. PE_3 is the in-zone competition winner, which gives the minimum "conscience" ED value of 1.2963, and its weight vector $\mathbf{W}_{31}$ is updated by equation (2) as shown in Table 3-5.

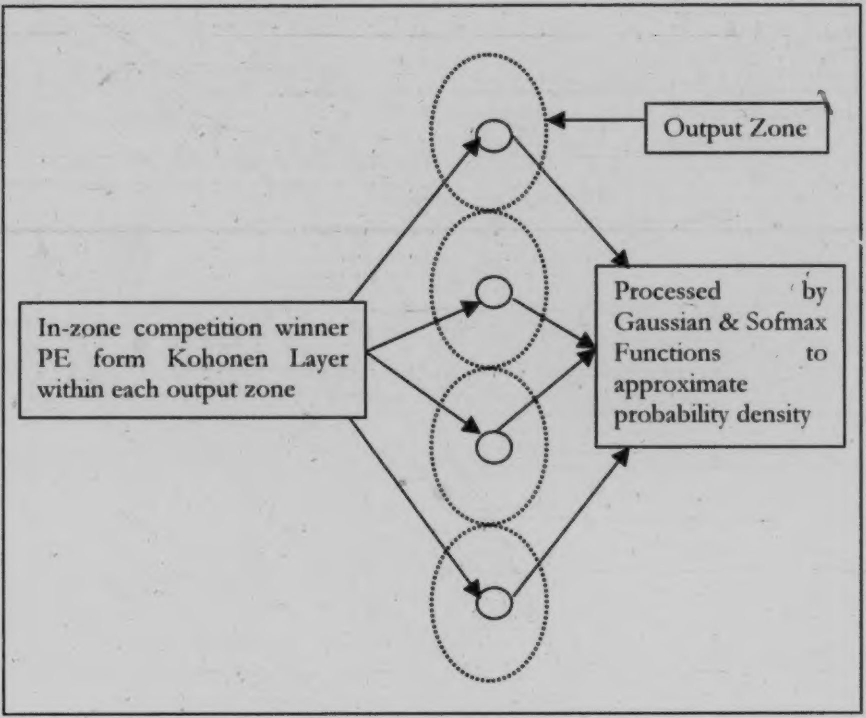

Output Zone
In-zone competition winner PE form Kohonen Layer within each output zone
Processed by Gaussian & Sofmax Functions to approximate probability density

Figure 3-2: Operations at Bayesian Layer in Recall

Table 3-5: Updating Weight Vectors in Second Leaning Stage

Input Vector X_1 (1)	PE_1 W_{11} (2)	PE_2 W_{21} (3)	PE_3 W_{31}' (4)
0.3	0.7232	0.0844	0.2564
0.2	0.6772	0.9129	0.2119
0.7	0.8780	0.4422	0.7541
0.8	0.6307	0.8774	0.8039
0.4	0.7523	0.7261	0.3994
0.5	0.8575	0.0439	0.4622
0.5	0.2088	0.4567	0.4762
0.9	0.6310	0.4542	0.7235
0.8	0.6065	0.2413	0.8276
0.1	0.8804	0.9188	0.2594
0.4	0.8232	0.7465	0.3555
0.6	0.0485	0.1158	0.6309

The above learning process will iterate through all the training records for a sufficient number of runs. Notice that during the learning process the Repulsion Rate, Attraction Rate, the conscience value are dynamically updated to calibrate the weight vectors.

In-Zone Competition Strategy at Kohonen Layer

The in-zone competition at the Kohonen layer occurring in the recall stage differs from that occurring in the learning stage. Once adequate training is complete, the PINN is capable of mapping the input onto the output. At the Kohonen layer, for one output zone, the PE that has the shortest global Euclidean distance (no conscience value) between its weight vector and the input vector is declared to be an in-zone winner PE. Only the in-zone winner PE advances to the Bayesian layer.

Unlike the GRNN/PNN, which takes the average of PE's Euclidean distance within one output zone as the parameter to pass forward (Specht 1988), PINN takes the minimum of the PE's Euclidean distance within one output zone as the parameter to pass forward. The reason for the difference is that in GRNN/PNN, each PE corresponds to one training record, and different numbers of PEs lie in different output zones. In the proposed PINN, the PE does not match the training record, and an equal number of PEs dwell in each output zones and work together in the Kohonen layer of PINN to generalize the underlying patterns within the training records by implementing LVQ.

PDF Approximator at Bayesian Layer

As illustrated in Figure 3- 2, at the Bayesian layer each output zone only contains the winner PE from the in-zone competition at the Kohonen Layer in the recall stage. The main components at the Bayesian layer are a kernel function and a Softmax Activation function that are used to approximate the probability density of one input vector being within each output zone in the steps as follow:

1. The square of Euclidean distance value of the winner PE from each zone is passed into the kernel function, which is the Gaussian function of Bayesian methods in statistics as described in Specht (1988) and shown in equation (3). If the number of

output zones is N, then for each input vector Xj, the kernel function is evaluated for N times and output one "q" value for each zone.

$$q_i = e^{[-(\mathbf{W}_{ij}-\mathbf{X}_j)^T(\mathbf{W}_{ij}-\mathbf{X}_j)/2\sigma^2]} \qquad (3)$$

where: i =1,2,3,…N, N is the number of output zones;

Xj is one input vector fed into PINN at the input layer;

$(\mathbf{W}_{ij}-\mathbf{X}_j)^T(\mathbf{W}_{ij}-\mathbf{X}_j)$ is the square of the Euclidean distance value between the input vector $\mathbf{X}_j$ and the winner PE's weight vector $\mathbf{W}_{ij}$ in the output zone i, i =1, 2,3,…,N.

σ is a smoothing factor, and is the only adjustable parameter of the Gaussian function (3) and controls the shape of the probability density function. The greater σ, the more dispersed the probability density graph. σ is critical to PINN's predicting capability and can be determined through iterative adjustments. In regard to industrial productivity application, σ should fall in the range between 0.8 and 1.2.

The Softmax Activation function (Sarle, 1997) as shown in equation (4) makes the sum of the calculation results (q values) from (3) equal to one, so that the final output from the Bayesian layer can be interpreted as posterior probabilities ("p" values).

$$p_i = \frac{q_i}{\sum_{i=1}^{N} q_i} \qquad (4)$$

where: q_i is the output value from Gaussian function (3) for output zone output zone i, i = 1, 2,3,…,N.

N is the number of output zones.

Outputs at Output Layer

At the output layer, the probability distribution predicted by the PINN is presented in the form of a Probability Density Function graph, which portrays the uncertainty of the output value.

In addition to the predicted distribution, PINN calculates two point prediction values:

1) Mode value: the median of the output zone or sub-range that has the greatest probability.
2) Weighted Average Value: the sum product of the median and the probability of each output zone. The user should treat this point-value prediction carefully by checking the probability density function graph.

Sample of Recall Processing

A sample PINN recall calculation is given in the following sections for illustration:

Suppose that the NN is trained and ready to recall the output for an input vector. As shown in Figure 3- 1, the dimension of input vector is 12, and the output range ([0-4]) is divided into 4 zones, i.e. [0-1], [1-2], [2-3], [3-4]. Each zone contains 3 processing elements.

Table 3-6 lists the scaled input vector $\mathbf{X}_1$ and the weight vectors of the 3 processing elements in zone 1.

Table 3-6: Trained PINN Ready to Recall for A Given Input Vector

Input Vector $\mathbf{X}_1$ (1)	PE_1 $\mathbf{W}_{11}$ (2)	PE_2 $\mathbf{W}_{21}$ (3)	PE_3 $\mathbf{W}_{31}$ (4)
0.5	0.8668	0.9681	0.6722
0.6	0.3437	0.4966	0.1627
0.4	0.6202	0.0093	0.5633
0.9	0.2472	0.2213	0.6362
0.4	0.3469	0.6834	0.4307
0.7	0.7326	0.4165	0.5033
0.8	0.6803	0.9864	0.8412
0.1	0.1366	0.4152	0.7824
0.3	0.0517	0.5621	0.8916
0.6	0.9558	0.5190	0.7239
0.9	0.9264	0.5150	0.3458
0.6	0.7467	0.5772	0.0914

The Euclidean distance (ED) values between the weight vectors and the input vector are calculated as ED_{11} =0.9513, ED_{21} = 1.1670, ED_{31} = 1.3249.

At output zone 1, processing element PE_1 with a minimum ED value (0.9513) is the in-zone competition winner, and proceeds to the Bayesian layer.

Suppose the winner PEs from the other 3 zones are also determined in the similar manner and proceed to the Bayesian layer. Table 3-7 lists their Euclidean distance values and outputs from the Gaussian function and the Softmax function.

Table 3-7: Recall Calculations at Bayesian Layer

Values for Each Zone (1)	Zone 1 (2)	Zone 2 (3)	Zone 3 (4)	Zone 4 (5)
Median	0.500	1.500	2.500	3.500
Winner PE's ED	0.951	2.567	1.843	2.922
Gaussian Output q:	0.636	0.037	0.183	0.014
Softmax Output p:	0.731	0.043	0.210	0.016

From Table 3-7, the probability of output being within zone 1 is 0.731, hence the mode output value is found to be the median of zone 1, i.e. 0.5. The weighted average output value is obtained by calculating the sum-product of the Softmax output (p values) and median of each output zone, i.e.

$$0.5\times0.731 + 1.5\times0.043 + 2.5\times0.210 + 3.5\times0.016 = 1.011.$$

IMPLEMENTATION OF THE PINN MODEL

Computer software based on the PINN model is developed for learning and testing in the environment of MS Access 97 and Visual Basic for Applications. Historical piping productivity data of 66 projects resulting in 119 records of a construction company was collected and compiled into NN input data for three labor-intensive activities, i.e. pipe installation, pipe welding and pipe hydro-testing. In the following sections, pipe installation is used to illustrate the testing and validation of the PINN model.

The PINN model for "pipe installation" has a total number of 81 input nodes. (The input factors and a sample data are shown in Table 3-1 and 2). The output range is divided into 20 output zones with an equal width of 0.72. 10 processing elements are assigned to each output zone. The attraction rate and repulsion rate are both equal to 0.06. The smoothing factor of the kernel function is equal to 0.8.

One hundred one records are used for PINN learning, while 18 records are reserved to test the calibrated model. The learning process takes 300 iterations.

Validation of the PINN Model on Testing Data

The testing of the calibrated network on the 18 unseen records was summarized in Figure 3- 3. Measured against the actual output values of the test data, for the mode value, the average absolute error is 0.57, and the maximum absolute error is 2.02; for the weighted average value, the average absolute error is 0.75, and the maximum absolute error is 2.23. Considering a wide output range of about 15, the error is reasonable and acceptable.

To compare the PINN model with a back propagation neural network, the same training records were used to train a three layer feed forward back propagation neural network, which has 81 input nodes at the input layer, 40 hidden nodes at the middle layer and 1 output node at the output layer. The training parameters are learning rate equal to 0.8, momentum rate equal to 0.4; the transfer function is symmetric logistic function. After training was completed, the testing set of 18 records previously used to test PINN was fed into the model. The testing results of the back propagation NN model comparing with that of the PINN are shown in Figure 3- 3. From Figure 3- 3, it is observed that the PINN model outperforms the back propagation NN model in terms of point prediction accuracy, coming closer to the actual output values.

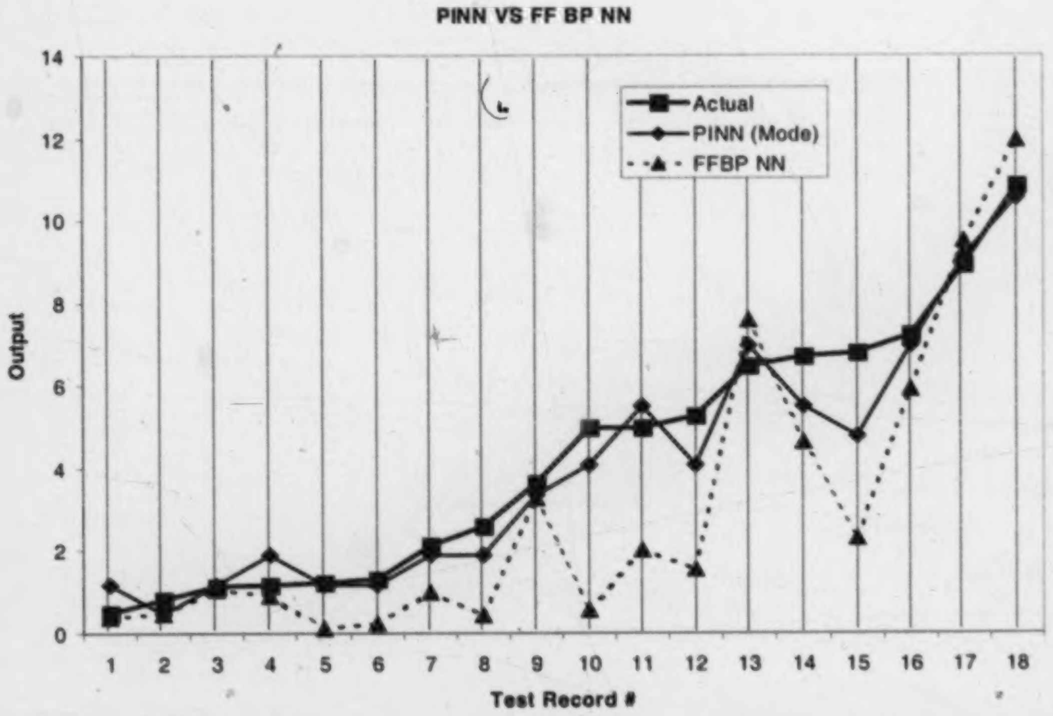
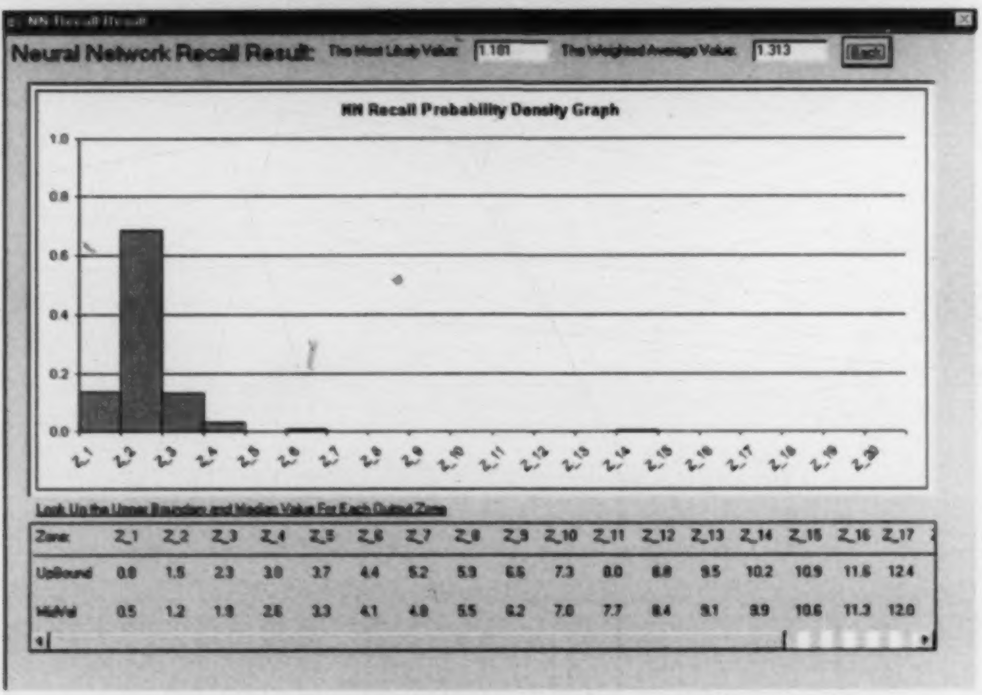


Figure 3-3: Comparison of PINN and Back Propagation NN


NN Recall Result
Neural Network Recall Result:
Back
NN Recall Probability Density Graph
1.0
0.8
0.6
0.4
0.2
0.0
Z_1
Z_2
Z_3
Z_4
Z_5
Z_6
Z_7
Z_8
Z_9
Z_10
Z_11
Z_12
Z_13
Z_14
Z_15
Z_16
Z_17
Z_18
Z_19
Z_20
Look Up the Upper Boundary and Median Value For Each Output Zone

The Most Likely Value: 1.181

The Weighted Average Value: 1.313

Zone:	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_10	Z_11	Z_12	Z_13	Z_14	Z_15	Z_16	Z_17
UpBound	0.8	1.5	2.3	3.0	3.7	4.4	5.2	5.9	6.6	7.3	8.0	8.8	9.5	10.2	10.9	11.6	12.4
MidVal	0.5	1.2	1.9	2.6	3.3	4.1	4.8	5.5	6.2	7.0	7.7	8.4	9.1	9.9	10.6	11.3	12.0

Figure 3-4: PINN Output for the Base Case Scenario

Sensitivity Analysis of the PINN Model

A recall program based on the trained PINN model for pipe installation was developed so as to validate its effectiveness and accuracy in the context of the application domain. "What if" scenarios are tested on the NN model by changing some input factors in order to understand the impact of such changes on the output values. The response of the NN model is compared against that of an experienced estimator at the participating construction company for the purpose of model validation.

The base case scenario is taken from one testing record. The actual difficulty multiplier for this scenario is 1.24. The mode value predicted by the PINN model is 1.181, giving an absolute error of (|1.181 - 1.24 | = 0.059); and the weighted average value is 1.313, giving an absolute error of (|1.313 - 1.24| = 0.073). Figure 3- 4 shows the predicted probability function or distribution, the chance of output falling into zone 2 ([0.8~1.5], median = 1.181) is 69%.

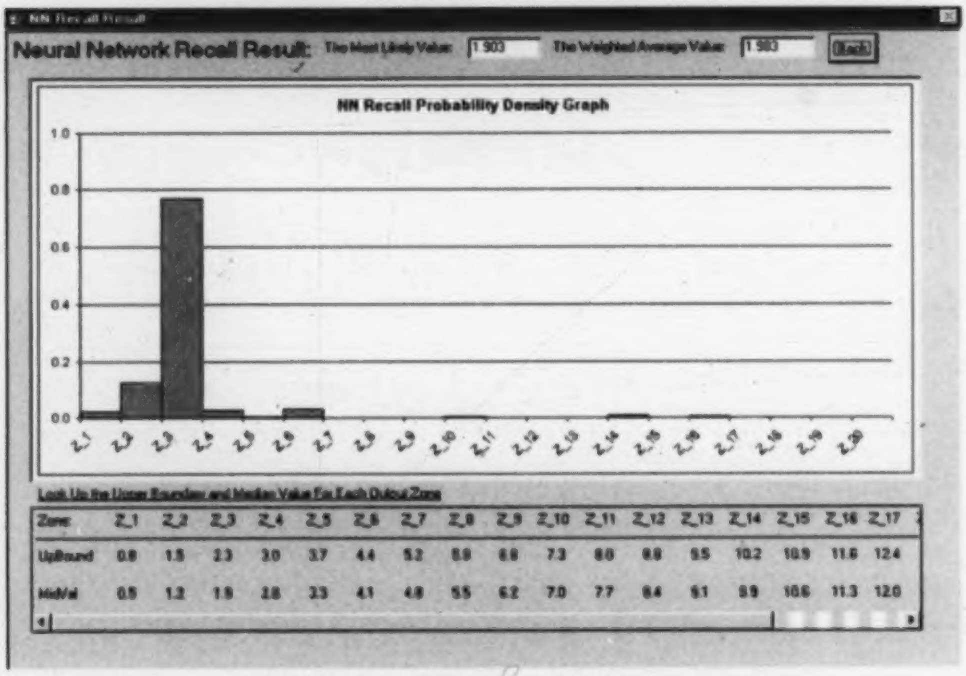
In the following validation tests, the actual values remain unknown, so the responses of the estimator based on personal experiences and common senses serve as a benchmark to measure the performance of the PINN model. The estimator responds with a trend or direction instead of a precise number because there are so other input factors to take into account.

SCENARIO 1

The location of pipe installation is a major consideration when the experienced estimator determines the pipe installation productivity. The installation location for the base case scenario is "Piping within a fabrication shop", what if the location is changed to "Operating plant installation on the site"? The experienced estimator responds by increasing the difficulty multiplier to a certain extent to reflect the unfavorable job conditions. Response of the PINN model is shown in Figure 3- 5. It is observed the mode value increases to 1.903 and the weighted average value increases to 1.983; the chance of the output value falling into zone 3 ([1.5~2.3], median = 1.903) increases from 13% in the base case scenario to 78%. The PINN has taken the same direction as the estimator in the decision process for this scenario.

SCENARIO 2

In the base case scenario, the job is done 100% in the winter in Alberta. What if

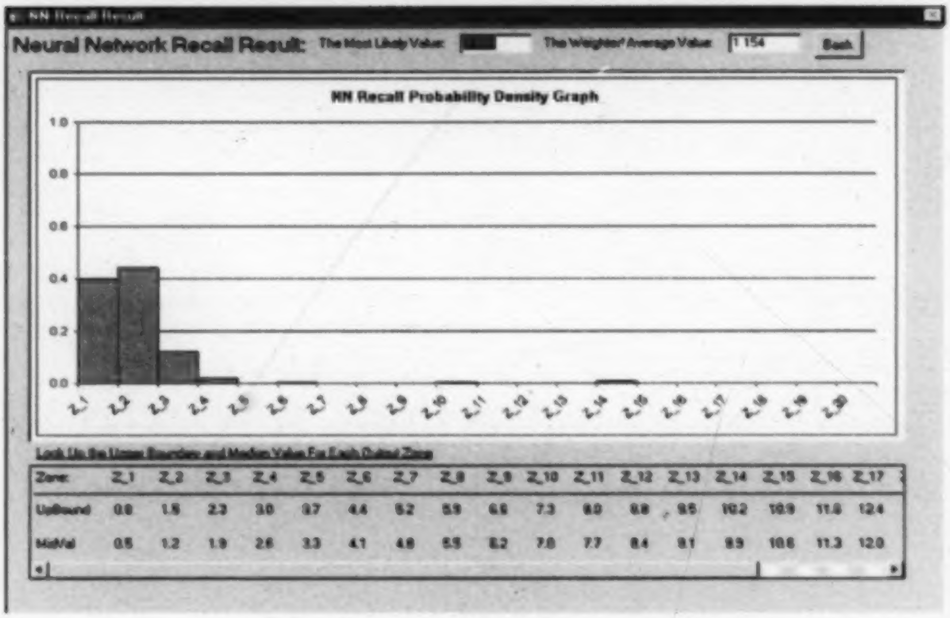

NN Recall Result
Neural Network Recall Result:
The Most Likely Value:
1.903
The Weighted Average Value:
1.983
Back
NN Recall Probability Density Graph
1.0
0.8
0.6
0.4
0.2
0.0
Z_1
Z_2
Z_3
Z_4
Z_5
Z_6
Z_7
Z_8
Z_9
Z_10
Z_11
Z_12
Z_13
Z_14
Z_15
Z_16
Z_17
Z_18
Z_19
Z_20

Look Up the Upper Boundary and Median Value For Each Output Zone

Zone:	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_10	Z_11	Z_12	Z_13	Z_14	Z_15	Z_16	Z_17
UpBound	0.8	1.5	2.3	3.0	3.7	4.4	5.2	5.9	6.6	7.3	8.0	8.8	9.5	10.2	10.9	11.6	12.4
MidVal	0.5	1.2	1.9	2.6	3.3	4.1	4.8	5.5	6.2	7.0	7.7	8.4	9.1	9.9	10.6	11.3	12.0

Figure 3- 5: PINN Output for Scenario 1

the job is done 100% in the summer? The estimator anticipates a reduction for the difficulty multiplier, which means an increase of productivity level. Response of the PINN model is shown in Figure 3- 6. It is observed the mode value remains 1.181, however, the weighted average value decreases to 1.154. The chance of the output value falling into zone 2 ([0.8~1.5], median = 1.181) decreases significantly from 69% to 44%, while the chance of the output value falling into zone 1 ([0.2~0.8], median = 0.5) increases significantly from 15% in the base case scenario to 40%. Again the PINN chooses the similar course of action as the estimator in the decision process for this scenario.


NN Recall Result
Neural Network Recall Result:
The Most Likely Value:
Back
NN Recall Probability Density Graph
1.0
0.8
0.6
0.4
0.2
0.0
Z_1
Z_2
Z_3
Z_4
Z_5
Z_6
Z_7
Z_8
Z_9
Z_10
Z_11
Z_12
Z_13
Z_14
Z_15
Z_16
Z_17
Z_18
Z_19
Z_20
Look Up the Upper Boundary and Median Value For Each Output Zone

The Weighted Average Value: 1.154

Zone:	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_10	Z_11	Z_12	Z_13	Z_14	Z_15	Z_16	Z_17
UpBound	0.8	1.5	2.3	3.0	3.7	4.4	5.2	5.9	6.6	7.3	8.0	8.8	9.5	10.2	10.9	11.6	12.4
MidVal	0.5	1.2	1.9	2.6	3.3	4.1	4.8	5.5	6.2	7.0	7.7	8.4	9.1	9.9	10.6	11.3	12.0

Figure 3-6: PINN Output for Scenario 2

CONCLUSIONS

The PINN model creates a meaningful representation of a complex, real-life situation in the problem domain and is effective in dealing with high dimensional input-output mapping with multiple influential factors in a probabilistic approach. The application of the PINN model in industrial labor production rate estimating helps the estimator choose a course of action by giving a better understanding of the project information available and the possible outcomes that could occur. Because the probability density of each output zone is provided, the predicted distribution and point-prediction values give the estimator much more confidence in the predicted result. In combination of the personal experiences and preferences labor production rate for a new project can be determined.

REFERENCES

Chao L.C., and Skibniewski, M.J. (1993). "Estimating Construction Productivity: Neural-Network-Based Approach." Journal of Computing in Civil Engineering, ASCE, 8(2), 234-215

Hermann, R. and Lu, M. (1999) "Application of Neural Networks in Industrial Estimating", Proceedings of the 27th Canadian Society of Civil Engineers Annual Conference, Edmonton, AB, 15-25.

Knowles, P. (1997). "Predicting Labor Productivity Using Neural Networks." Masters of Science Thesis, University of Alberta, Edmonton, AB.

Karshenas, S. and Feng, X.(1992) "Application of Neural Networks in Earthmoving Equipment Production Estimating." Computing in Civil Engineering, Proceeding of Eighth Conference, Dallas, Texas, 841-847

Kohonen, T. (1995). "Self-Organizing Maps", Springer Series in Information Sciences, Springer, London, U.K.

Moselhi, O., Hegazy, T., and Fazio, P. (1990) "Neural Network as Tools in Construction." Journal of Construction Engineering and Management, ASCE, 117(4), 606-623

Murtaza, M.B., and Fisher, D.J. (1993), "Neuromex: Neural Network System for Modular Construction Decision Making", Journal Computing in Civil Engineering, ASCE, 8(2), 221-233.

Portas, J., and AbouRizk, S.M. (1997), "Neural Network Model For Estimating Construction Productivity." J. of Constr. Engrg. & Mgmt., ASCE, 123(4), 399-410.

Sarle, W.S., ed. (1997), Neural Network FAQ, part 1 of 7: Introduction, periodic posting to the Usenet newsgroup comp.ai.neural-nets, URL: ftp://ftp.sas.com/pub/neural/FAQ.html

Specht, D. F.(1988). "Probabilistic Neural Networks for Classification, Mappings, or Associative Memory." IEEE International Conference on Neural Networks, 1988, 1, 525-532.

Specht, D. F. (1991) "General Regression Neural Networks." IEEE Transactions on Neural Networks, Nov. 1991, 2, 525-576.

APPENDIX:

DESINO'S METHOD TO CALCULATE "CONSCIENCE" VALUE

"Conscience" Euclidean Distance is defined as (5):

$$Di' = Di + Ci \qquad (5)$$

Where:

Di is Euclidean distance,

and Ci is conscience value (6):

$$C_i = cf \times D \times (n \times wf - 1) \qquad (6)$$

Where:

D is the maximum Euclidean distance out of the global competition in the previous supervised learning stage,

cf is a Conscience Factor, which is initially set between 0 and 1 by the user,

and n is the number of PEs per output zone,

wf is defined as "Win Frequency", and the initial estimate of the Win Frequency value (wf_0) is set to the reciprocal of PE Number per Output Zone for all the PEs, i.e. 1/n.

With NN learning ongoing, both Conscience Factor (cf) and Frequency Estimate (fe) are reduced gradually until it approaches 0 at the end of learning.

During the unsupervised learning stage, for the in-zone winner PE, its wf value is updated as (7):

$$wf' = (1 - wf_0) \times wf + wf_0 \qquad (7)$$

For the in-zone loser PEs, their wf values are updated as (8):

$$wf' = (1 - wf_0) \times wf \qquad (8)$$

Basically, the above formulas are intended to increase the wf values for the winner PE and hence increase its conscience value so that the winner PE will have less chances to win again than the other loser PEs in the following learning iterations.

CHAPTER 4: SENSITIVITY ANALYSIS OF NEURAL NETWORKS IN SPOOL FABRICATION PRODUCTIVITY STUDIES[1]

INTRODUCTION

Artificial neural networks (NN) mimic the cognitive learning process in the human brain, and deal effectively with ill-structured problems, in which the algorithms required to solve them cannot be given in a precise and explicit fashion, or the data for a particular problem are either not complete or cannot be specified precisely (Widman et. al., 1989). NN has been found to be capable of performing parallel computations on different tasks, such as pattern recognition, linear optimization, speech recognition, and prediction (Mukherjee and Deshpande 1995).

---

[1] A version of this chapter has been submitted for publication. ASCE, Journal of Computing in Civil Engineering.

In recent years, Back Propagation NN (BPNN) has been researched and applied as a convenient decision-support tool in a variety of application areas in civil engineering, including modular construction decision making (Murtaza and Fisher, 1993), structural analysis (Flood and Katim, 1994), estimating construction productivity (Portas and AbouRizk, 1997), mode choice analysis of freight transport market (Sayed and Razavi, 1999), construction markup estimating (Li et al 1999), measuring organizational effectiveness (Sinha and McKim, 2000), and predicting settlement during tunneling (Shi, 2000). The special learning algorithms of BPNN are capable of performing high dimensional, non-linear input-output mapping and extracting hidden patterns and predictive information from observing the learning examples.

However, learning algorithms such as BPNN do not attempt to infer causality, hence, classification or prediction is based on blind correlation of new examples with previously analyzed examples, without giving information on the effect of each input parameter or influencing variable upon the predicted output variable. In the reported NN applications, model validation has thus far relied upon measuring accuracy of the calibrated network to an independent testing data set that are hidden from the neural network in learning. The model's sensitivity to changes in its parameters is generally probed by testing the response of a mature network on various input scenarios. In short, a NN model functions like a "black box" package, giving no clue on (1) how the answers or model outputs are obtained; (2) how the input parameters affect the output.

Widman et. al (1989) pointed out that the credibility of an AI program frequently depends on its ability to explain its conclusions. Lack of interpretability is a pitfall of the neural network models recognized by many and has inhibited NN from achieving its full potential in real-world applications. Dhar and Stein (1997) argued that because NN algorithms such as the back-propagation NN are non-linear, high dimensional functional equations featuring parallel distributed data processing, it is hard to explicitly interpret which parameters cause what behavior in the NN model. While mathematical and operational methods do exist for the analysis of neural networks, the methods are fairly involved, and are less than satisfying because of their theoretical assumptions. They stated that "unlike most statistical methods, it can be difficult to say, even in general, which variables are significant in what respect." (Dhar, et.al. 1997)

Our research intends to address the identified issue by concentrating on sensitivity analysis of BPNN. Similar to regression analysis, the sensitivity of an NN input parameter could be expressed as the first-order partial derivative between an NN output variable and the input parameter. In the "Literature Review" section, we briefly introduce several related methods for knowledge explanation and factor analysis of NN found in literature. In the section entitled "BPNN Algorithm and Input Sensitivity", the back-propagation NN algorithm is described first, followed by the derivation of mathematical relationships between an NN output variable and NN input space in light of both normalized data and original data. The following section "BPNN vs. Regression Analysis" discusses the difference between BPNN and regression analysis of statistics and demonstrates the sophistication and superiority of BPNN over regression analysis in a case study based on a small data set. Next, statistical analysis of input sensitivity based on Monte Carlo simulation is described in the section entitled "Statistical Analysis of Input Sensitivity" to understand the rationale of BPNN's reasoning and the effectiveness of model implementation in a probabilistic fashion. In the "Industrial Application" section, the new approach is applied to estimate the labor productivity of spool fabrication in an industrial setting, and important aspects of the application including problem definition, factor identification, data collection, and the testing results based on real data set are discussed and presented.

LITERATURE REVIEW

Li et al (1999) realized the inability of BPNN to provide explanations on its output negatively affects the user-acceptance of BPNN. They investigated the use of KT-1 method for automatically extracting rules from a mature BPNN in an attempt to explain why and how BPNN makes a particular recommendation in developing a decision support tool to estimate the construction markup. KT-1 method is a heuristic approach to generating confirming/disconfirming rules from each hidden or output node based on the weighted connections and the threshold value of each node, and is constrained by the complexity of network structure. As pointed out by the authors, such automatic rule-extraction systems as KT-1 cannot warrant a fully informative explanation facility because KT-1 lacks the associative knowledge (i.e. common sense, professional knowledge etc.) in its rule-extracting process (Li, et. al., 1999).

Sinha and McKim (2000) utilized BPNN to measure organizational effectiveness of construction firms. They have applied statistic analysis methods, such as Principal Component Analysis (PCA), stepwise regression and correlation analysis on a BPNN model in an attempt to identify the dominant factors that influence the target output variable, and further to reduce the dimension of input space. However, the theoretical underpinning of such statistical techniques requires careful study of their applicability in solving real problems. For instance, lacking an awareness of the assumptions of least squares regression (normality, homoscedasticity, independence of errors, and linearity) may cause the misuse of regression and correlation analysis (Levine et. al, 1998); use of PCA, which assumes linear relationships between variables, might bias the selection of determinant factors by excluding those that have non-linear relationships with the target output variable (Refenes, et. al.,1995).

Alternative approaches to factor analysis include using auto-associative back-propagation neural networks to perform non-linear dimension reduction and sensitivity analysis (Carrol and Ruppert, 1988). The neural network has one hidden layer with k hidden processing elements, where k is less than the dimension of input space n. The output space is a replica of the input space. Analogous to the rationale of PCA, this method is to compress data by representing many variables by a few components: if we can reproduce the input space using k ($k < n$) hidden processing elements without loss of information, then the activation values of the k processing elements in the hidden layer will compute the first k principal components at the input space, under appropriate conditions. One pitfall of this method observed by Refenes et. al (1995) is that the stochastic nature of the data-generating process at the neural network input space may cause high variance in the analysis results. It is also noted that the output variable is excluded from the auto-associative neural network analysis, hence, such analysis of input parameters or influencing factors does not take into account the relationships between the input parameters and the output variable.

Explanations on the importance of input parameters can also be obtained by examining the weights of a mature network so as to characterize the strengths of the relationship between inputs and outputs. Knowles and AbouRizk (1997) added up the absolute value of weights from one input node to every hidden processing element in a trained BPNN model with only one hidden layer for estimating pipe installation productivity. The total weight value of an input node may indicate the intensity of the connection from the input node to the hidden layer of the network; the higher the sum value, the more significant the input parameter is. Although this heuristic approach is straightforward to understand and easy to use, the results may be unstable or inaccurate due to the fact that it fails to take into account the connections between the hidden layer and output layer. In modeling the behavioral mode choice of the U.S. freight transport market, Sayed and Razavi (1999) combined the learning ability of BPNN and the transparent nature of fuzzy logic in order to explain the knowledge contained in a BPNN model, which is stored in the form of a weight matrix that is hard to interpret. The neurofuzzy model facilitates the selection of significant variables that affect the output and displays the stored knowledge in terms of fuzzy linguistic rules (Sayed and Razavi, 1999).

Based on the above methods found in literature, the effect of each input parameter on the output variable in terms of magnitude and direction still remains unknown, i.e. the input sensitivity. In the following section the algorithmic aspects of the BPNN model are studied in order to define the input sensitivity in an exact mathematical term.

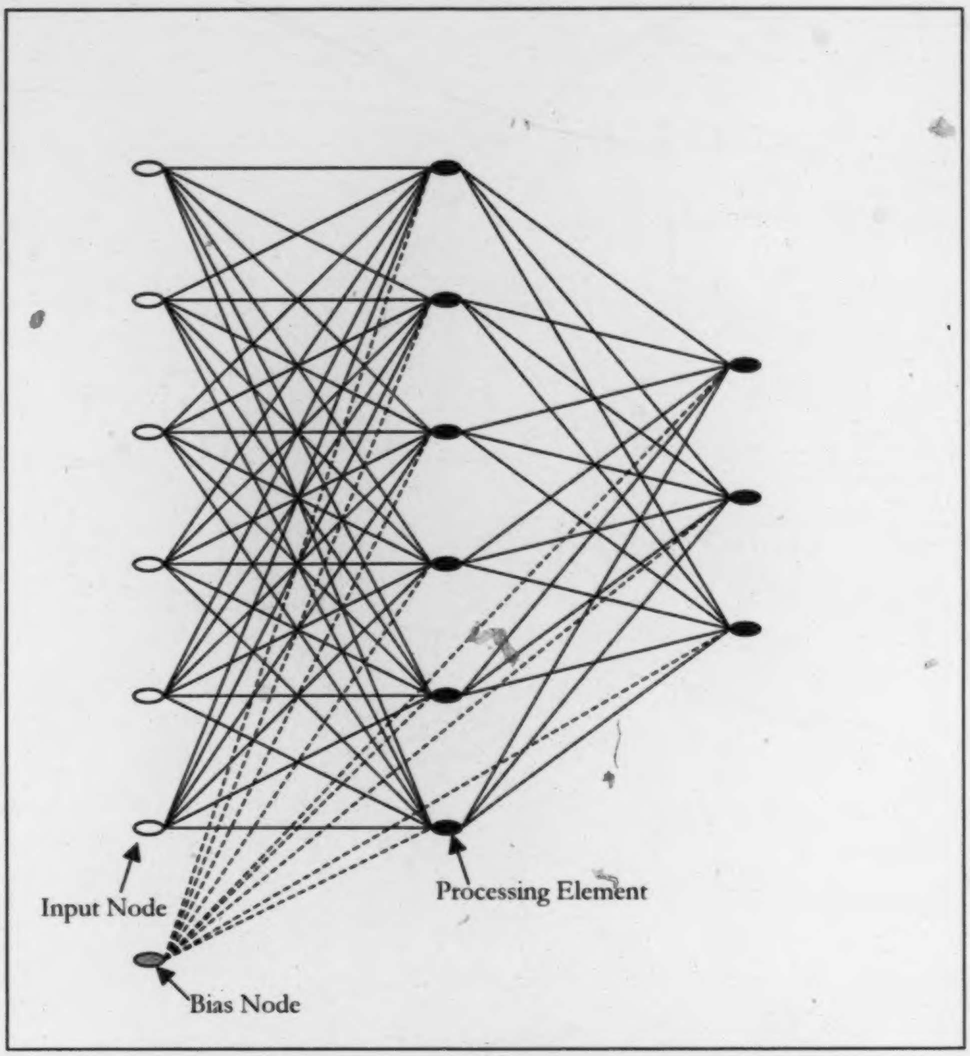

Input Node
Processing Element
Bias Node

Figure 4-1: Structure of Back-Propagation NN Model

BPNN ALGORITHM AND INPUT SENSITIVITY

From the biological perspective, BPNN is originally proposed as an AI model to simulate the cognitive learning process in human brain, in which millions of neurons are interconnected and interact with one and another through complex electrochemical reactions and signal processing. An artificial NN model such as BPNN is merely an over-simplified representation of the real NN in terms of mechanism and structure. A typical BPNN has a multi-layer structure. Each layer contains a number of processing elements (PE) or nodes, which are fully interconnected between layers (Figure 4-1). The intensity of connection between two processing elements is represented using a weight. Put into the perspective of mathematics, BPNN is essentially a gradient-decent optimization algorithm to search for the optima in a high-dimensional weight space with the objective of minimizing the global error between NN output values and actual output values. An iterative weight-adjusting scheme is used to modify the weights of all the connections in the NN structure in a stepwise fashion.

BPNN Algorithm

The basic formulae to describe signal processing of a PE in BPNN are simply as:

$$S_c = \sum N_i \cdot W_{ic} \tag{1}$$

$$N_c = \frac{1}{1+e^{-S_c}} \tag{2}$$

Where:

Subscript c stands for a processing element in the hidden layer or output layer of BPNN;

Subscript i is the node index at the previous layer of BPNN;

W_{ic} stands for a weight value between node i and node c;

S stands for the input signal to a node;

N stands for the output signal from a node.

Equation (1) shows that a processing element receives a weighted linear combination of input signals from the previous layer. Equation (2) is Sigmoid (logistic) function and is the most commonly used transfer (squashing) function in BPNN, through which a processing element transforms the input signal into an output signal. Note that a bias node with constant input value –1 from the previous layer is also connected to a processing element and involved in the calculation, representing the activation threshold of a processing element (Figure 4-1).

The digital signals flow through the BPNN following (1) and (2) from layer to layer until the output layer is reached.

The global error (E) of the BPNN optimization search is expressed in (3):

$$E = \frac{1}{2T}\sum_{i=1}^{T}(N_i - D_i)^2 \qquad (3)$$

Where:

N stands for the output signal from BPNN;

D stands for the target value;

Subscript i is the index of records in the training data set and T stands for the size of training data set.

The weight of the BPNN is adjusted using the delta rule to move to the opposite direction of $\frac{\partial E}{\partial W_{pc}}$ as (4):

$$\Delta W_{pc} = -\lambda \cdot \frac{\partial E}{\partial W_{pc}} = -\lambda \cdot \frac{\partial E}{\partial S_c} \cdot \frac{dS_c}{dW_{pc}} = -\lambda \cdot N_p \cdot \frac{\partial E}{\partial S_c} \qquad (4)$$

Where:

λ is a gain ratio in (0,1), also called learning rate, which sets the pace of BPNN learning;

Subscript p stands for a processing element in the previous layer of the network;

Subscript c stands for a processing element in the current layer of the network.

For a processing element at the output layer of BPNN,

$$\frac{\partial E}{\partial S_c} = N_c \cdot (1 - N_c) \cdot (N_c - D_c) \qquad (5)$$

For a processing element at the hidden layer of BPNN,

$$\frac{\partial E}{\partial S_c} = N_c \cdot (1 - N_c) \cdot \sum_{n=1}^{J} \frac{\partial E}{\partial S_n} \cdot W_{cn} \qquad (6)$$

In (6), subscript n is the index of processing elements in the next layer, and J is the total number of processing elements in the next layer.

Usually, a momentum term is added to the weight adjusting scheme to take into account the weight change in the previous step as shown in (7).

$$\Delta W_{pc} = -\lambda \cdot N_p \cdot \frac{\partial E}{\partial S_c} + \mu \cdot \Delta W_{pc}' \qquad (7)$$

$\Delta W_{pc}'$ is the weight change in the previous step, and μ is the momentum ratio which is usually less than λ.

BPNN adjusts weights following (7) by observing the training data set repeatedly, until the global error E is reduced to an accepted level to declare the BPNN to be trained.

The BPNN should have at least three layers: the input layer, one hidden layer, and the output layer. The three-layer-structured BPNN with Sigmoid transfer functions has been found by many to be adequate in solving non-linear optimization problems. In the following sections, we use the three-layer Sigmoidal BPNN to illustrate the mathematical inferences of input sensitivity for simplicity of representation. However, interested readers can readily extend the derived input sensitivity to BPNN with more complex structures and other transfer functions.

Input Sensitivity Based on Normalized Data

Based on the BPNN algorithm presented in previous sections, we can sort out the relationships between an output variable and an input parameter to define the input sensitivity of BPNN in an exact mathematical term.

Notations used in the following mathematical formulae are listed as below:

- Subscript p stands for a node in the Previous layer of the network.

- Subscript c stands for a node in the Current layer of the network; C stands for the total number of nodes in the current layer.

- Subscript n stands for a node in the Next layer of the network.

- W_{ij} stands for the weight of connection between node i and node j.

- S stands for the input signal to a node.

- N stands for the output signal from a node.

If the current node is an input node (in the first layer of the network), S is the normalized input data in range (0,1).

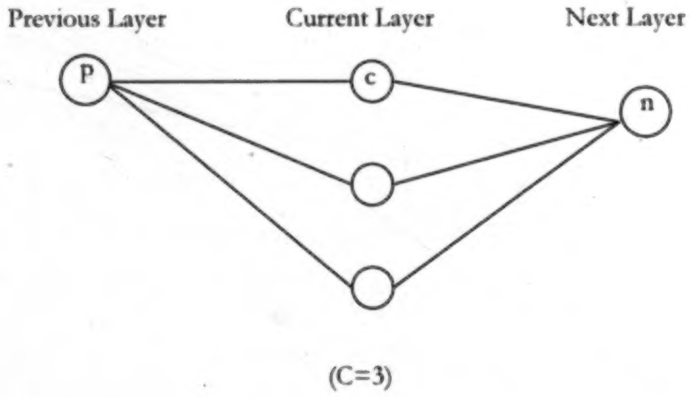


Figure 4-2: Illustration for Node and Layer Representations

$$N_c = S_c \qquad (8)$$

If the current node is not an input node (in a hidden layer or an output layer),

$$S_c = N_p \cdot W_{pc} + \sum_{i \neq p} N_i \cdot W_{ic} \qquad (9)$$

$$\frac{\partial S_c}{\partial N_p} = W_{pc} \qquad (10)$$

Note that equation (9) is the same as (1) except that (9) distinguishes one node or processing element p from others at the previous layer. The relationship between the output signal N_c and the input signal S_c is defined in (2), from which, we have:

$$\frac{dN_c}{dS_c} = N_c(1 - N_c) \qquad (11)$$

As shown in Figure 4-2, in the three-layer BPNN, node p is an input node in the input layer, node c is a hidden node in the middle layer, and node n is an output node at the output layer. The focus of BPNN sensitive analysis is on investigating the first-order partial derivative of the output signal from node n (N_n) over the input signal to node p (S_p). By (8), we have $S_p = N_p$

$$\frac{\partial N_n}{\partial S_p} = \frac{\partial N_n}{\partial N_p} = \frac{\partial N_n}{\partial N_c} \cdot \frac{\partial N_c}{\partial N_p} = \left(\frac{dN_n}{dS_n} \cdot \frac{\partial S_n}{\partial N_c} \right) \cdot \left(\frac{dN_c}{dS_c} \cdot \frac{\partial S_c}{\partial N_p} \right) \qquad (12)$$

from (10), we know,

$$\frac{\partial S_c}{\partial N_p} = W_{pc}, \text{ and } \frac{\partial S_n}{\partial N_c} = W_{cn}$$

From (11), we know,

$$\frac{dN_c}{dS_c} = N_c(1 - N_c), \text{ and } \frac{dN_n}{dS_n} = N_n(1 - N_n)$$

So, (12) can be expressed as:

$$\frac{\partial N_n}{\partial S_p} = N_n(1 - N_n)W_{cn} \cdot N_c(1 - N_c)W_{pc} \tag{13}$$

Because more than one processing element exists in the current layer (hidden layer), assume, the number of processing elements at layer c is C. A general form of input sensitivity for BPNN is then expressed as (14):

$$\frac{\partial N_n}{\partial S_p} = \sum_{i=1}^{C} W_{pc_i} W_{c_i n} \cdot N_{c_i}(1 - N_{c_i})N_n(1 - N_n) \tag{14}$$

Input Sensitivity based on Original Data

In deriving (14), we assume all data including inputs and outputs has already been normalized in the range (0,1). From the perspective of real applications, usually it is convenient and straightforward to probe the sensitivity of BPNN based on the original or raw data instead of scaled data.

Various linear or non-linear normalization methods can be used to transform raw data. Shi (2000) reviewed the established data transformation methods for BPNN and proposed a new one called "distribution transformation", which fits statistical distributions to raw input data and utilizes the resultant Cumulative Density Functions (CDF) to scale inputs to [0,1]. The theoretical underpinning of such transformation is relatively weak as pointed out in Shi, 2000, but such transformation does complicate the application of NN. The conclusion about the superiority of "distribution transformation" scenario over the traditional linear transformation scenario is arrived at empirically based on independent experiments on each method. Due to a number of variable factors (such as learning rates, momentum) and stochastic phenomena (such as the initialization of network weights and the existence of multiple local optima in the searching space), the improvement of network performance may not be attributable or only partly attributable to the input transformation methods.

The non-linear mapping capability of BPNN is mainly owing to the non-linear transfer functions in hidden and output PE s. According to our experiments on BPNN, a good selection of hidden layer structures and transfer functions based on trials generally results in improvement of BPNN's performance. Thus, we recommend using such robust, undistorted and simple data normalization methods as linear transformation to normalize both inputs and outputs in BPNN and satisfy the neural computation requirements. The simplicity of BPNN will be maintained without sacrificing its functionality, which can be further demonstrated from sensitivity analysis of BPNN in the following sections.

If we take into consideration the data normalization procedure, the simplest and most commonly used one is a linear process as follows:

$$N_p = \frac{(UB - LB)}{(MAX_p - MIN_p)} \cdot (S_p - MIN_p) + LB \qquad (15)$$

Where, UB is the upper bound of the normalized interval (LB,UB), and LB is the lower bound, for sigmoid transfer function, usually LB = 0, and UB = 1;

MAX_p is the maximum value in the data set corresponding to input node p or parameter p;

MIN_p is the minimum value in the data set corresponding to input node p or parameter p.

A formula similar to (15) is applied to normalize the output data, in order to match the output range of the transfer functions in BPNN, i.e. (0,1) for Sigmoid transfer functions. If we take N_n as the raw output data, there is a scale-back process involved at the output layer, which will cancel out the UB (1) and LB (0) in combination with the scaling process at the input layer. So we can arrive at a more general form of (14) based on linear normalization procedures as:

$$\frac{\partial N_n}{\partial S_p} = \frac{MAX_n - MIN_n}{MAX_p - MIN_p} \cdot \sum_{i=1}^{C} W_{pc_i} W_{c_i n} \cdot N_{c_i}(1 - N_{c_i}) \cdot N_n(1 - N_n) \qquad (16)$$

This slope or partial derivative is defined as absolute input sensitivity and represents the expected change in output variable N_n, per unit (1) change in input parameter S_p, holding the other input parameters constant. In a real-world problem, each input parameter may have different unit of measure, and hence various relevant range, which encompasses all values from the smallest to the largest used in training the model. Similar to regression analysis, it is important for BPNN to interpolate within the range rather than extrapolate beyond the range in order to make sensible predictions. For one input parameter ranging from 1 to 10000, one unit change is too small to be considered while for another input parameter ranging from 0.1 to 0.6, one unit change is too big to occur. Thus, it is more appropriate to use a relative one-unit (such as10% of relevant ranges) as the basic unit change in input parameters instead of an absolute one-unit (1). Through such transformation, the input sensitivity is undistorted and more meaningful in terms of comparing the effect of different input parameters upon the output variable. The relative input sensitivity is defined as (17):

$$\frac{\partial N_n}{\partial S_p}^{R} = \frac{MAX_n - MIN_n}{10} \cdot \sum_{i=1}^{C} W_{pc_i} W_{c_i n} \cdot N_{c_i}(1 - N_{c_i}) \cdot N_n(1 - N_n) \quad (17)$$

Equation (17) shows the relative input sensitivity $\frac{\partial N_n}{\partial S_p}^{R}$ based on 10% of relevant ranges i.e. 10% times (MAX_p - MIN_p). Note that the input sensitivity is independent of the relevant ranges of input parameters and represents the amount that output variable N_n changes (either positive or negative) for a particular unit change in the input parameter S_p, i.e. the 10% of its relevant range.

BPNN VS. REGRESSION ANALYSIS

The above sensitivity analysis of BPNN is analogous to the classic multiple regression analysis in statistics, which predicts the values of response or dependent variable based on the values of multiple explanatory or independent variable and can be defined as (18):

$$N_n = \beta_0 + \sum_{i=1}^{M} \frac{\partial N_n}{\partial S_{pi}} \cdot S_{pi} \tag{18}$$

Where β_0 is an intercept representing the average value of N_n when all the explanatory variables S_{pi} are equal to zero, i =1 to M, M is the total number of explanatory variables. $\frac{\partial N_n}{\partial S_{pi}}$ is a slope for the i^th^ explanatory variable and its definition is identical to the input sensitivity of BPNN as above. However, only by examining the difference between BPNN and regression analysis can the sophistication and superiority of BPNN over regression analysis be demonstrated, as discussed next.

An examination of Equation (16) indicates that the value of $\frac{\partial N_n}{\partial S_p}$ in BPNN is dependent on several factors:

1. The internal structure of BPNN, i.e. the number of hidden nodes and number of hidden layers.

2. The BPNN data set, i.e. the relevant range of each input parameter and output variable.

3. The weight values of BPNN, i.e. the intensity of connection among processing elements from the input layer to the output layer. This is actually the result of BPNN training, and hence dependent on the training data set.

4. The current input values loaded at the input nodes. From (1), (2), and (4), it is evident N_c and N_n are functions of the current input values at the input layer and the weight values of BPNN.

We can also observe that once a BPNN is trained on a data set, the first three factors (BPNN structure, weights and training data set) are fixed, so the sensitivity of an input parameter over an output variable is totally determined by the fourth factor, i.e. the current input values. If we treat the current input values as the coordinate values of an input point at the BPNN input space, the dimension of which is equal to the number of input parameters, we can conclude that, for a trained BPNN,

$$\frac{\partial N_n}{\partial S_p} = F(Input\_Point) \qquad (19)$$

Here, F stands for a function.

Indeed, BPNN performs a multiple linear regression analysis at each individual data point to fit a non-linear high dimension hyperplane to the training data set. The slope value along each dimension, along with the intercept value β_0, varies from data point to data point, in contrast with being constant in regression analysis. Simply put in two dimension space, BPNN is capable of fitting a flexible curve and all the observed data points fall on the line; while regression analysis can only approximate a straight line that strings up the data points with the minimum amount of deviation based on least square method.

Aside from above discussions, three other advantages of BPNN over regression analysis are worth mentioning:

(1) BPNN poses no theoretical constraints on data in contrast with the assumptions of least square regression and the required residual analysis in regression analysis (Levine et al, 1997).

(2) BPNN supports more than one output in input-output mapping in contrast with only one output in regression analysis.

(3) BPNN relaxes the requirements of data in terms of both quantity and quality in contrast with regression analysis. That means BPNN is capable of non-linear mapping with only a very limited quantity of observed data points and is tolerant of noisy data (inaccurate or incomplete data).

Table 4-1: Data Set for Testing BPNN and Regression Analysis

Input_1 (1)	Input_2 (2)	Input_3 (3)	Input_4 (4)	Output (5)
0.1	0.1	0.1	0.1	0.1
0.1	0.2	0.3	0.4	0.3
0.2	0.5	0.4	0.8	0.6
0.3	0.6	0.7	0.2	0.4
0.4	0.3	0.5	0.7	0.5
0.5	0.4	0.4	0.3	0.2
0.5	0.5	0.5	0.5	0.5
0.6	0.5	0.1	0.4	0.8
0.9	0.8	0.8	0.6	0.7
0.9	0.9	0.9	0.9	0.9

In order to illustrate the comparison of BPNN and regression analysis, we studied the input sensitivity of BPNN trained on an artificial data set with 4 inputs, 1 output and only 10 records as shown in Table 4-1. The BPNN model has four input parameters, 1 output variable, and one hidden layer with three hidden nodes, which is determined based on trials. The learning rate is 0.8 and the momentum is 0.4. Standard Error of the estimate in regression analysis is a measure of variation around the fitted line of regression and is calculated as a measurement of accuracy to compare the performances of two techniques. Standard Error is actually a slight variant of the global error term E in BPNN as (3). After achieving satisfactory training (standard error of the NN output is reduced to 0.00158), we calculated the partial derivative values of the output variable over each input parameter using (16) at various input points. The results, as shown in Table 4-2, indicate that for a specific input parameter, the slope value over the output variable varies with the input points. In order to analyze such variation, a Monte Carlo simulation is performed at the BPNN input space to observe the statistics of $\frac{\partial N_n}{\partial S_p}$ value for each input parameter. In each simulation run, an input point is randomly generated in the BPNN input space and triggers a BPNN recall process. A slope value of each input parameter over the output variable is calculated. If the number of simulation runs is large enough, we can assume we will traverse the entire BPNN input space by interpolating. A program in MS VB and Access is developed to perform the Monte Carlo simulation experiments for 1000 iterations. The resultant Probability Density Functions (PDF) of slope values for the four input parameters are shown in Figure 4-3 and the statistics are summarized in Table 4-3.

Table 4-2: Partial Derivative (Slope) ($\frac{\partial N_1}{\partial N_p}$) at Four Input Points at BPNN Input Space

Input Factor Index (p) (1)	Point (0.5,0.5,0.5,0.5) (2)	Point (0.1,0.1,0.1,0.1) (3)	Point (0.9,0.9,0.9,0.9) (4)	Point (0.2,0.4,0.6,0.8) (5)
1	-0.1594	-0.3057	0.2169	0.1115
2	0.8915	0.2614	0.3267	0.3947
3	-0.1685	-0.0398	-0.4169	0.3402
4	0.9974	0.4799	0.2878	0.1611

Table 4-3: Statistics of Partial Derivative (Slope) Values: ($\frac{\partial N_1}{\partial S_p}$)

Input Factor Index (p) (1)	Maximum (2)	Minimum (3)	Average (4)	Std. Dev. (5)	95% Confidence Interval (6)
1	0.9797	-1.0366	-0.0151	0.3863	-0.0390 ~ 0.0089
2	1.9502	0.0042	0.5769	0.4038	0.5518 ~ 0.6019
3	0.4833	-2.1053	-0.2678	0.5489	-0.3018 ~ -0.2338
4	2.2255	0.0031	0.6365	0.5064	0.6051 ~ 0.6679

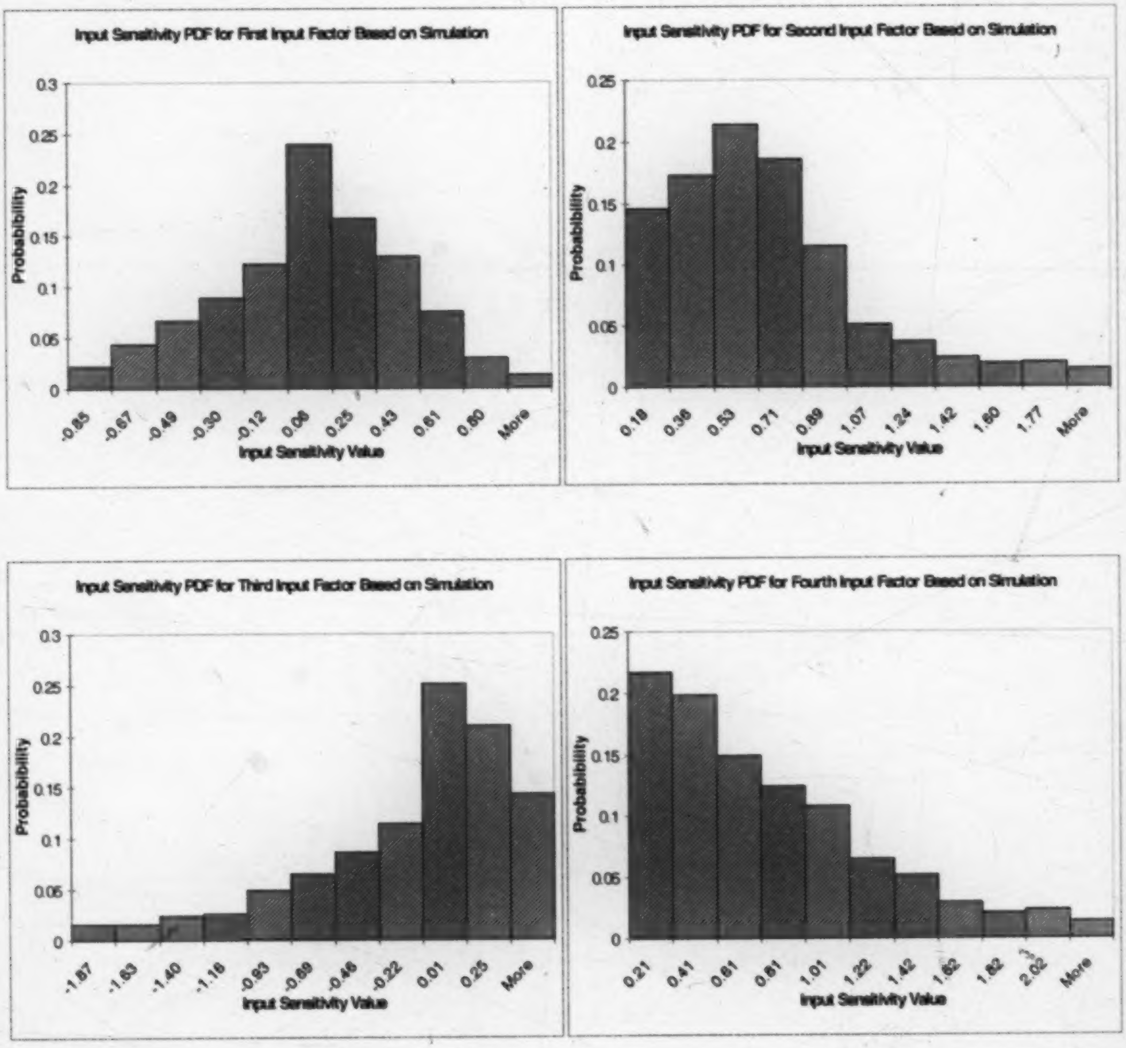


Figure 4-3: Distributions for Input Sensitivity

A regression analysis is conducted on the same data set of 10 observations in MS Excel. The results are that the slope of the first input parameter is 0.1217, the slope of the second input parameter is 1.0141, the slope of the third input parameter is minus 0.5925, and the slope of the fourth input parameter is 0.5509; the intercept is minus 0.03054. Note that those slope values are constants in contrast with distributions as obtained from BPNN. The standard error based on the outputs of regression analysis is as high as 0.1185 compared with merely 0.00158 of BPNN.

In short, BPNN outperforms regression analysis by a significant margin in our experiments, which agrees with the previous analysis and comparisons.

STATISTICAL ANALYSIS OF INPUT SENSITIVITY

The simulation results reveal distributions of slope data for BPNN, which take various shapes (Fig. 3). If the actual distribution of input sensitivity to be encountered in operations is available, comparison of the actual distribution with the corresponding Monte Carlo distribution obtained from BPNN can serve as an effective means for model validation. However, in most real BPNN applications quantitative information is unavailable to fit such actual distributions of input sensitivity due to the complexity of the engineering or management problems being solved. This is also the reason of choosing BPNN instead of other conventional mathematical models in the first place. An experienced domain expert may also have difficulty figuring out such distributions of input sensitivity on a subjective basis, because the decision process generally relies on assessment of the entire input scenario and there are so many interacting factors. The domain experts may share some common hunches about the probability of increasing or decreasing the output variable with a certain adjustment of an influencing factor. But the amount of adjustment is generally very subjective depending on the input scenario and personal experience and temperament.

Therefore, instead of fitting distributions, statistical analysis of simulation results involves calculating 5 percentiles of the slope variable for each input parameter, i.e. the 10th, 25th, 50th, 75th, and 90th. The input sensitivity of all input parameters is summarized and presented in a tornado-like graph as illustrated in Figure 4-4 for the piping fabrication labor productivity BPNN model. The horizontal axis represents the relative input sensitivity as determined by (17), i.e. output response (negative or positive) with a change of 10% relevant range in an input parameter. The vertical axis is the baseline corresponding to no output response or zero change in output. Five short vertical bars correspond to each input parameter representing respectively the five percentiles from left to right, reflecting the central trend, the spread, and the shape of the observed slope data distribution from simulation. The guidelines for interpreting the graph and simulation results are listed as below:

- The leftmost bar (10th percentile) being to the left of baseline represents that the chance of the slope value for the corresponding input parameter being positive is above 90%, or with increase of the input value the probability for the output value to increase is 90%.

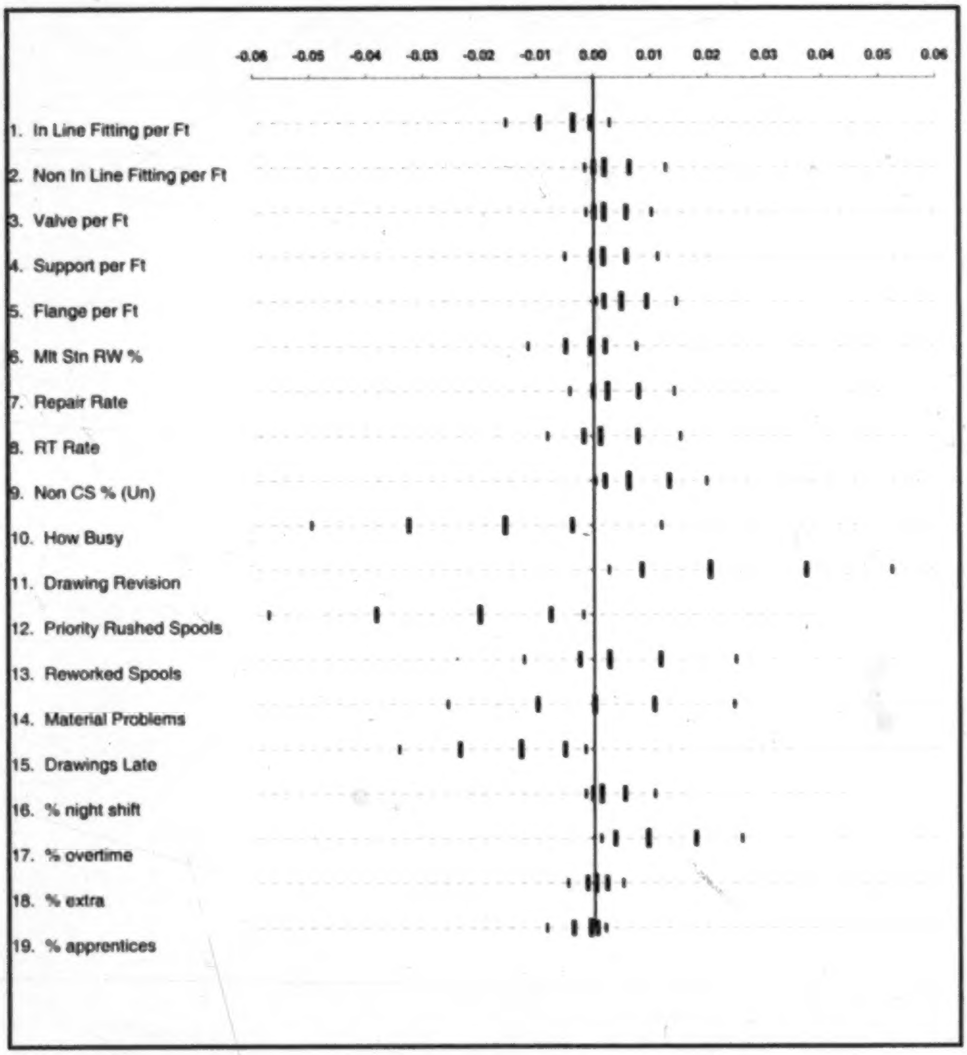


Figure 4-4: Sensitivity Analysis of Spool Fabrication BPNN Model

- The rightmost bar (90th percentile) being to the right of baseline represents that the chance of the slope value for the corresponding input parameter being negative is above 90%, or with increase of the input value the probability for the output value to decrease is 90%.

- The 25th and 75th percentiles can be explained in a similar manner as the 10th and 90th percentiles according to the relative positions of the corresponding bars to the baseline in the graph.

- The middle bar (50th percentile) riding on the baseline represents that the chance for the output variable to increase or decrease is 50%.

- An input parameter with a slope distribution clustering around the baseline has less effect on the output variable than that with a slope distribution distant from the baseline. Thus, the magnitude of input sensitivity can be inferred from observing the absolute values of percentiles as well.

- Note that the statistical descriptors (percentiles) are based on simulation samples rather than the entire population. However, the sample size is assumed to be large enough (10000 runs to draw Figure 4-4) to traverse the input space of BPNN, and the confidence interval estimates are rather tight, hence the statistical descriptors based on the samples can represent those for the population.

- The proposed sensitivity analysis method is of stochastic nature because of independent trials for BPNN training (such as initialization of network parameters, hidden layer structure, and local optima) and Monte Carlo process. If BPNN training is achieved and the simulation iteration is large enough, the results for most input parameters are stable in terms of direction and magnitude of input sensitivity, except for a couple of input parameters swapping sides with respect to the baseline from trial to trial. A semi-optimal BPNN model can be determined by selecting the trial in which input sensitivity of major input parameters makes sense or is agreed upon by the domain expert.

- In case that the sensitivity of one input parameter always takes the opposite direction in the tornado graph comparing against domain expert's experience or common sense, the definition and data collection procedures for the input parameter along with the data itself should be carefully reexamined for shortfalls before the input parameter is dropped out of BPNN analysis.

INDUSTRIAL APPLICATION

The sensitivity analysis of BPNN as described in the previous sections is applied to analyze a BPNN model for estimating labor production rate of pipe spool fabrication in the fabrication facility of PCL Industrial Constructors Inc, which is one of the largest and most modern pipe fabrication and module facilities in Western Canada.

Spool Fabrication Basics

A pipe spool is a portion of piping system consisting of various piping components, such as flanges, elbows, reducers, tees, supports, and pipe. These items are prefabricated into distinct assemblies that are later assembled together as part of an industrial plant or production skid/module. Such prefabrication is usually performed under controlled shop environment located away from the actual project site, which allows for better productivity and quality control, and hence cuts the field labor costs.

Major spool fabrication processes, such as cut, bevel, fit, weld, and handle sections of pipe and fittings, tends to be labor-intensive. Productivity data is collected for 63 projects completed from 1995 to 1999, during which period the technologies and machines for welding and cutting in the shop remain relatively stable. The productivity studies of spool fabrication is suitable to the unit-cost estimating method, in which labor production rates must be independent of equipment use and vary among projects only because of differences in labor productivity (Parker et. al., 1984).

Due to the variation in size, wall thickness and configuration of each individual spool, a special unitization scheme is utilized in the company to quantify the various work items uniformly into an abstract unit of measure called "Fabrication Unit" or "Unit" on the basis of weld inches of standard wall thickness pipe. Quantity of non-welding work items such as cutting, beveling, handling pipe and fittings, installing supports are also converted into "Units" by applying corresponding empirical factors in the scheme.

Factor Identification and Data Collection

The labor hours per fabrication unit become the focus of investigation, which ranges from 0.1 MH/Unit to 0.5 MH/Unit in the collected historical data. The unit labor hours fluctuate from job to job due to a number of quantitative and qualitative factors, including the complexity of spool configuration, the material components in fabrication, the stringency of quality control, spool drawing quality, the amounts of night shift and overtime, extra work, crew experience etc. The environmental effects and management factors are not considered as significant factors because of the controlled shop environment and consistent policy and personnel of management during the period of investigation. 19 input parameters are identified as listed in Table 4-4.

Table 4-4: Input Factors of Spool Fabrication Labor Productivity

ID (1)	NN Input Factor (2)	Data Source (3)	Remarks (4)
1	In Line Fitting (pcs) per Foot of Pipe in Spool	Material Track. Sys.	A ratio indicating the average length of pipe sections in spool
2	Non In Line Fitting (pcs) per Foot of Pipe in Spool	Material Track. Sys.	A ratio indicating complexity of spool configuration
3	Valve (pcs) per Foot of Pipe in Spool	Material Track. Sys.	A ratio indicating complexity of spool configuration
4	Support (pcs) per Foot of Pipe in Spool	Material Track. Sys.	A ratio indicating complexity of spool configuration
5	Flange (pcs) per Foot of Pipe in Spool	Material Track. Sys.	A ratio indicating complexity of spool configuration
6	Multi-Station Roll Weld Inches / Total Roll Weld Inches	Weld Track. Sys.	Multi-Station Roll Weld requires extra handling between weld stations
7	Repair Rate	Weld Track. Sys.	An index of crew's proficiency
8	Radiography Test Requirement	Weld Track. Sys.	An index of quality control stringency by specs.
9	Non CS Units / Total Units	Weld Track. & Material Track. Sys.	Non CS component in fabrication requires extra care in storage, handling and welding
10	Shop Work Load	Questionnaire	A 5-point rating based on shop workload in units and no. of concurrent jobs indicating how busy the shop was.
11	Drawing Revision Rate	Questionnaire	A 5-point rating based on percent of revised spool drawings indicating drawing quality
12	Priority Rushed Spools	Questionnaire	A 5-point rating based on percent of rushed spool due to client priority indicating shop work schedules.
13	Rework Spools	Questionnaire	A 5-point rating based on percent of reworked spools due to drawing errors and quality defects
14	Material Shortage Problems	Questionnaire	A 5-point rating on efficiency of material supply
15	Late Drawing Issues	Questionnaire	A 5-point rating based on percent of late spool drawing issuance by client that impacts fabrication
16	Night Shift MHs / Total MHs	Payroll Sys.	Night Shift affects labor productivity
17	Over Time MHs / Total MHs	Payroll Sys.	Over Time affects labor productivity
18	Extra Work MHs / Total MHs	Payroll Sys.	Extra Work affects labor productivity
19	Apprenticeship MHs / Total MHs	Payroll Sys.	Welder qualification system affects labor productivity: Apprentice vs. Journeyman

Data is collected from the company's various transaction systems including labor cost tracking system, weld tracking system, payroll system, material tracking system. In order to ease the burden of data gathering and ensure high quality of data, a historical project data warehouse is custom-developed using Microsoft Access and VBA to integrate raw data from different transaction systems and automate the validation of raw data and the calculation of productivity information. Because data is unavailable in current transaction systems of the company for such factors as the drawing revision rate, late drawing issuance, material shortage problems, quantity of reworked spools, quantity of rushed spools due to priority, shop work load data, a questionnaire survey is carefully designed and conducted with the support of the company management. The key personnel involved in the projects including shop superintendents, project managers and coordinators, QC staff, and welding foremen are interviewed to help recall some facts and gather the needed information.

BPNN Training and Sensitivity Analysis

A total number of 70 records are compiled and used to train a BPNN model with 19 input nodes at the input layer corresponding to 19 input parameters, 19 hidden nodes at the middle layer, and 1 output node at the output layer that is the unit labor hours. The number of hidden nodes can be determined based on trials; BPNN learning is found to be unsusceptible when the number of hidden nodes is close to the number of input nodes. The learning rate is 0.4, the momentum is 0.1, and sigmoid transfer functions are used in hidden and output nodes. After satisfactory training (standard error of the output is 0.00143), the Monte Carlo based sensitivity analysis is performed on the matured network for 10000 simulation runs. Note that Equation (17) is used to determine the input sensitivity, which is based on the change of 10% of relevant input range.

Several independent trials from BPNN training to the sensitivity analysis are conducted on the same data set. The best trial, in which the input sensitivity of most factors follows the same trends, as determined by experienced domain experts, is shown in Figure 4-4. An examination of Figure 4-4 reveals the relationships between the influencing factors and the fabrication productivity, which are generalized by BPNN through observing historical project data in the past 5 years. For example, factor 1 is about in line fitting pieces per foot of pipe in spool, which indicates the average length of pipe sections in spool. According to our domain experts, in line fittings, such as unions, couplings, swages, reducer etc are used to connect pipe sections in a straight line without turns or branches. Thus, the more in line fitting pieces in spools, the more small sections of pipe in spools, and the easier to handle the work. From Figure 4-4, BPNN determines the chances to decrease labor hours per unit with the increase of this ratio are about 78% and agrees with the trend identified by domain experts. Factors 2 to 5 are four ratios indicating the complexity of spool configuration. By our domain experts, the higher such ratios, the more complex the spools' configuration, and the tougher to fabricate the spools. From Figure 4-4, the dominant trends of the four ratios are all on the plus side, which matches the experience of our domain experts. It is also observed from Figure 4-4 that factor 18 (extra work percentage) is relatively tightly enveloped around the baseline, which indicates that extra work is not as dominant as other factors in contributing to the variance in unit labor rates. The explanation can be partly attributed to the fact that the amount of extra work more directly impacts the efficiency of administration or management than the productivity of crew on the shop floor. Other input factors can be interpreted and validated in a similar manner, and are not elaborated further due to space limit.

Model Testing and Validation

In particular, the effect of material type of spool fabrication on the labor productivity is tested based on the BPNN model, because material type (carbon steel, stainless steel, aluminum etc.) is a major consideration of an industrial estimator in adjusting unit labor hours of spool fabrication. The labor production rate of non-carbon steel fabrication is empirically 1.5 times the rate of carbon steel in company's business guideline. 24 records in the data set with 0% non-carbon steel component (100% carbon steel fabrication) are selected as testing records. In the next step, for each testing record, only the input parameter of non-carbon steel component is changed from 0% to 100% with other parameters intact. Those testing records are fed to the network and let BPNN recall the output, i.e. the unit labor rates for non-carbon steel fabrication. The output from BPNN is compared against the original output of each record, i.e. the unit labor rate for carbon steel fabrication. Based on the test results in Figure 4-5, BPNN increases the unit labor hours on 75% of the records; the amount of decrease for 5 records, i.e. No. 1, 2, 5, 6, 9, is relatively small comparing with the amount of increase for others. If the sample size is large enough, the percentage should come close to about 90%, as observed from Figure 4-4 for Factor 9. On average, the ratio of non-carbon steel labor rate over carbon steel labor rate is 1.4, which is close to 1.5 in the guideline.

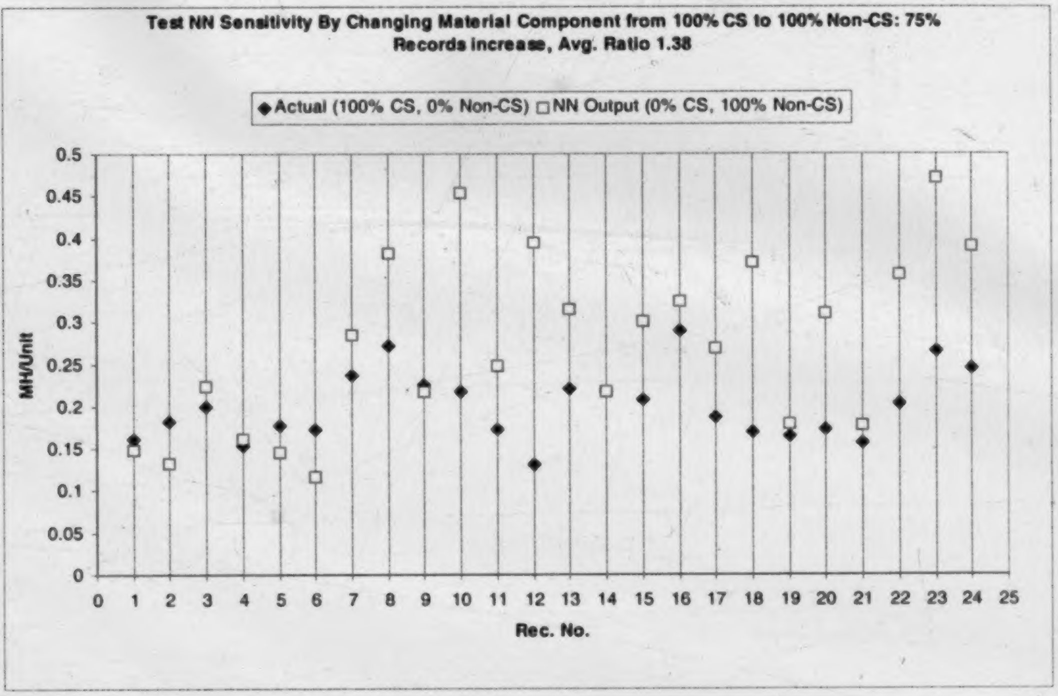


Figure 4-5: Testing Sensitivity of BPNN to Material Type

Note that the guideline gives an average number (1.5) in consideration of material type only, while BPNN is able to figure different numbers for different scenarios taking into account 19 relevant factors. In short, a BPNN-based decision support tool will be more sophisticated and intelligent than the traditional business guideline.

CONCLUSIONS

The model validation approach of BPNN based on the proposed sensitivity analysis is superior to the conventional validation approach of testing the mature network with an independent data set, in that such sensitivity analysis enables the modeler to understand the rationale of BPNN's reasoning and have a pre-knowledge about the effectiveness of model implementation in a probabilistic fashion.

The insight into the BPNN model gained from the proposed sensitivity analysis method gives the user more confidence in the BPNN's prediction, hence facilitates the implementation of BPNN-based decision support tools. The success of our industrial application in estimating labor productivity of spool fabrication exceeded our initial expectations. Not only does this new method prove to be effective in addressing problem domains in which BPNN has been applied, but also it potentially makes BPNN appealing to new engineering or business applications.

REFERENCES

Carrol, R.J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*, Chapman & Hall, New York, NY.

Dhar, V. and Stein, R. (1997). *Intelligent Decision Support Systems: The Science of Knowledge Work*. Upper Saddle River, Prentice-Hall, Inc., New Jersey.

Flood, I., and Kartam, N. (1994). "Neural networks in civil engineering: systems and applications." *J. Constr. Engrg. And Mgmt.*, ASCE, 124(1), 18-23.

Knowles, P. (1997). *Predicting Labor Productivity Using Neural Networks.* Masters of Science Thesis, University of Alberta, Edmonton, AB.

Levine, D. M., Berenson, M. L., and Stephan, D. (1998). *Statistics for Managers using Microsoft EXCEL*, Prentice-Hall, Inc., Upper Saddle Rive, New Jersey.

Li, Y., Shen, L. Y., and Love, P.E.D. (1999). "ANN-based mark-up estimation system with self-explanatory capabilities." *Journal of Construction Engineering and Management*, ASCE, 125(3), 185-189.

Mukherjee, A., and Deshpande, J.M. (1995), "Modeling initial design process using artificial neural networks", *Journal Computing in Civil Engineering*, ASCE, 9(3), 194-200.

Murtaza, M.B., and Fisher, D.J. (1993), "Neuromex: Neural Network System for Modular Construction Decision Making", *Journal Computing in Civil Engineering*, ASCE, 8(2), 221-233.

Shi, J. (2000). "Reducing prediction error by transforming input data for neural networks", Journal Computing in Civil Engineering, ASCE, 14(2), 109-115.

Sinha, S. K. and McKim, R.A. (2000). "Artificial neural network for measuring organizational effectiveness.", Journal Computing in Civil Engineering, ASCE, 14(1), 9-14.

Parker, A.D., Barrie, D. S., and Snyder, R. M. (1984), Planning and Estimating Heavy Construction, McGraw-Hill, Inc., New York, NY.Portas, J., and AbouRizk, S.M.

(1997). Neural Network Model For Estimating Construction Productivity. *J. of Constr. Engrg. & Mgmt.*, ASCE, 123(4), 399-410.

Refenes, A.N., Zaprnis, A.D., Connor, J.T. and Bunn, D.W. (1995). "Neural Networks in Investment Management". *Intelligent Systems for Finance and Business*, S. Goonatilake, and P. Treleaven, eds., John Wiley & Sons Ltd, , Chichester, England, 177-209.

Sayed, T., and Razavi, A. (1999). "Comparison of Neural and Conventional Approaches to Mode Choice Analysis" *Journal of Computing in Civil Engineering*, ASCE, 14(1), 23-30.

Widman, L.E., and Loparo, K.A (1989). "Artificial intelligence, Simulation, and modeling: a critical survey", *Artificial intelligence, simulation, and modeling*. L.E.Widman, K.A. Loparo, and N.R. Nielsen, eds., John Wiley & Sons Ltd, New York, NY, 1-45.

Chapter 5: Conclusion and Recommendation

SUMMARY OF THESIS WORK

In conjunction with a major industrial contractor of Canada, the thesis research conducted case studies on the theoretical basis and practical considerations for measuring and analyzing labor productivity in industrial construction. Two important activities of process piping were investigated: pipe installation in the field and spool fabrication in the fabrication shop. The primary objective of research is developing ANN-based estimating tools to offer estimators valuable information about labor productivity in bidding new jobs, because estimating labor productivity is one of the most difficult aspects of preparing an estimate, or a control budget based on the estimate for labor-intensive activities in industrial construction. Artificial neural networks are capable of sorting out hidden patterns and extracting predictive information from complex data sets, and were proven to be effective in both uncertainty analysis and sensitivity analysis of construction labor productivity in the research. The thesis research has addressed: (1) how to quantify labor productivity in industrial construction from a contractor's point of view; (2) how to measure actual labor productivity in industrial construction based upon on-site control practices; and (3) how to utilize Artificial Neural Networks (ANN) to analyze the variability of actual labor production rates and the sensitivity of identified influencing factors.

Productivity Studies and Data Collection

The thesis research reviewed current estimating practices as applied to the involved company and generalized special methods utilized in practice for the quantification and measurement of labor productivity in industrial construction. The input factors that cause the variability in the productivity for studied activities were identified through literature review and consultation with experienced domain experts at the involved company. With the support of the company's management, two data warehouses were custom-developed for field pipe installation and shop spool fabrication respectively to integrate the corporate management systems of estimating, production resources planning, quality control, and labor cost control. It should be mentioned that questionnaire surveys were carefully designed and conducted to collect some qualitative and descriptive information that is not obtainable from the company's reporting and accounting systems. Experienced superintendents, project managers and estimators of the involved company were interviewed to help recall some facts and gather the needed information. The data warehouses provide solid platform of integrated historical data from which to validate novel ANN models and develop ANN-based tools for productivity analysis.

Probabilistic Neural Network Modeling

The thesis research derived a probabilistic neural network classification model called the Probability Inference Neural Network (PINN), which is based on the same concepts as those of the Learning Vector Quantization (LVQ) method combined with a probabilistic approach. The PINN model was intended to overcome limitations of other neural network models and was developed for predicting labor production rates for industrial construction. The thesis presented and explained the topology and algorithm of the PINN model in details. Portable computer software was developed to implement the training, testing and recall for PINN. PINN was tested on real historical productivity data at the involved company to analyze the degree-of-difficulty factor of field pipe installation productivity and compared to the classic feed forward back propagation neural network model; this showed marked improvement in performance and accuracy. The PINN model creates a meaningful representation of a complex, real-life situation in the problem domain and in general is effective in dealing with high dimensional input-output mapping with multiple influential factors in a probabilistic approach. The application of the PINN model in industrial labor production rate estimating gives an estimator a better understanding of the project information available and the possible outcomes that could occur. Because the response of PINN is in the form of a probability density function (distribution) at the output range, an estimator will be able to decide on the degree-of-difficulty factor for a future scenario by combining the PINN's recommendation with personal judgment.

Sensitivity Analysis of Back Propagation Neural Networks

Validation of a NN model has thus far relied upon measuring accuracy of the calibrated network to an independent testing data set that are hidden from the neural network in learning. A NN model's sensitivity to changes in its parameters is generally probed by testing the response of a mature network on various input scenarios. The thesis research also investigated the classic back propagation NN algorithm to study the effect of each input parameter or influencing variable upon the predicted output variable. The input sensitivity of back propagation NN is defined in exact mathematical terms in light of both normalized data and raw data. The difference between back propagation NN and regression analysis of statistics is discussed and the sophistication and superiority of back propagation NN over regression analysis is further demonstrated in a case study based on a small data set. In addition, statistical analysis of input sensitivity based on Monte Carlo simulation enables the modeler to understand the rationale of back propagation NN reasoning and have pre-knowledge about the effectiveness of model implementation in a probabilistic fashion. The sensitivity analysis of back propagation NN was successfully applied to analyze the labor production rate of pipe spool fabrication at the involved company. Important aspects of the application including problem definition, factor identification, data collection, and model testing based on real data were discussed and presented in the thesis. The model validation approach of back propagation NN based on the proposed sensitivity analysis is superior to the conventional validation approach, in which the mature network is tested with an independent data set and the model's sensitivity is probed through observing the output with respect to changes in input based on a limited number of scenarios. The insight into the back propagation NN model gained from the proposed sensitivity analysis method gives the user more confidence in the back propagation NN's prediction, hence facilitates the implementation of back propagation NN -based decision support tools. Not only does this new method prove to be effective in addressing problem domains in which back propagation NN has been applied, but also it potentially makes back propagation NN appealing to new engineering or business applications.

Conclusion

The problems addressed in the thesis research were identified through investigating the current estimating practices in industry and understanding the real concerns of industry professionals. Emerging computer modeling techniques such as data warehouses and ANN were researched from an academic perspective and implemented in industry to meet with the challenges. The proposed novel ANN models and developed decision support tools were validated using real data from industry and successfully applied to assist estimators in deciding on labor production rates for new jobs. The experiences and lessons learned from the successful, productive and mutually beneficial collaboration between academia and industry throughout the thesis research will potentially serve as a model to guide other university-industry joint research projects in the future.

FUTURE RECOMMENDATIONS

There are a number of issues that need to be addressed in greater detail in the future.

Quantification of Textual / Descriptive data

Three input data types are used to define NN input factors in the thesis, i.e. "Raw", "Rank", and "Binary". "Raw" is used simply for quantitative input factors, like general expense ratios, winter construction percentages, or quantities of work. "Rank" is used to convert subjective factors, like crew ability ratings, into numeric format. And "Binary" is used to group textual or descriptive factors into numeric formats like material type and project definition. It should be noted an input factor of the "Raw" or "Rank" type corresponds to one input node at the input layer, while an input factor of the "Binary" type corresponds to a number of input nodes depending on the number of groups for the factor. "Binary" data type satisfies the computing requirements of neural networks for converting textual or descriptive data, however, some disadvantages associated with "Binary" data may affect the performance and sensitivity analysis of neural networks.

First, increased dimension of NN input space caused by "Binary" data type increases the complexity of network structure and the quantity of network parameters. Based on experimentations and observations, the PINN model is not very susceptible to the increase of the NN input space dimension, however back propagation NN does suffer in terms of learning time and generalization ability with the increase of input dimensionality. The generalization ability is not guaranteed to improve, but chances are very high that the learning time will increase considerably.

Secondly, the input sensitivity of back propagation neural networks is defined for each NN input node. A change for an input factor of "Binary" data type entails changes in more than one input nodes of NN. Thus the input sensitivity for an input factor of "Binary" data type must take into account the combination effect of involved input nodes. What input nodes are involved depends on how the change is made. For example, suppose four different material types are considered, corresponding to four NN input nodes, a change from type 1 (1000) to type 2 (0100) triggers changes in the first and second NN input nodes; while a change from type 1 (1000) to type 3 (0010) triggers changes in the first and third NN input nodes. Note that the input sensitivity of back propagation NN for one input node is not a constant value but a distribution, such combination effect makes it difficult to explain the sensitivity of an "Binary" type factor.

Fortunately, there is no such "Binary" type factor in spool fabrication productivity analysis where the sensitivity analysis was tested. For the field pipe installation productivity analysis, an experiment was conducted to treat such "Binary" factors such as material type and project type as "Rank" factors. Various groups in each factor were ranked on a 5-point scale by their relative difficulty based on the judgment of domain expert such that a unit increase in the corresponding NN input node could represent the increase of degree-of-difficulty factors. The results of the experiment are satisfactory and the input sensitivity follows the correct direction for most factors. However, the drawback of such a heuristic method is that sometimes even the domain experts found it hard or impossible to weight the relative difficulty and rank each group in a factor in a sensible way.

In short, more sophisticated methods such as fuzzy set theory may be researched and introduced into NN to convert textual or descriptive factors into numeric formats.

Optimization of NN Structure

NN structure mainly concerns with the middle layers, for instance, the number of hidden layers and number of hidden nodes in each for a BP NN; the number of processing elements assigned to each output zone at Kohonen layer and the setup of output zones for a PINN model. The determination of NN structure relies heavily on a trial-and-error based process, in which comparing the NN's outputs with actual outputs on an independent testing data set serves as a yardstick for justifying the structure. Such optimization of NN structure tends to be hampered due to factors such as the stochastic processes involved in NN learning, the existence of multiple local optima in search of the NN internal parameters, noise within learning data set, values of learning rates and types of data transfer functions

One appealing solution is to obtain the actual distribution of input sensitivity for key input factors (if not all) to be encountered in operations. Matching the corresponding Monte Carlo distributions obtained from BP NN to the actual distributions can serve as an effective means for optimization of the NN structure, in addition to validation of the NN model as discussed in Chapter 4. Hence, gathering quantitative information to fit actual distributions of input sensitivity could be included as part of data collection for NN applications in the future if both data and resources are available.

Sensitivity Analysis of PINN Model

In the thesis, the PINN model's sensitivity to changes in its parameters is still probed by testing the response of a mature network on various input scenarios. One approach that have been tried is to take advantage of the sensitivity analysis method for BP NN as proposed in the thesis to infer the input sensitivity for a PINN model under the following conditions:

1. The BP NN and PINN are trained with the same learning data set and tested with the same testing data set, and both models are satisfactorily trained;
2. And the point-value predictions of the BP NN and those of the PINN model for the testing data set are very close.

Thus, it can be assumed that the two models would "think alike" and have common input sensitivity for a particular input factor.

Table 5-1 shows the results of five testing records based on the training data for spool fabrication productivity after two models have satisfied the above conditions.

Table 5-1: PINN vs. BP NN

BPNN	PINN (Mode)
0.239	0.205
0.156	0.145
0.221	0.235
0.157	0.145
0.286	0.265

It is noted that the first condition is not hard to satisfy, but the second condition may not be readily met. The BP NN and PINN may require to be trained repetitively using different structures and learning parameters in order to satisfy both conditions.

In the future, it would be perfect if an independent approach could be found to explain the input sensitivity of the PINN model analytically.

FINAL REMARKS

The applications of ANN in estimating labor productivity of industrial construction prove that ANN is effective in addressing the complexity and requirements in the problem domain. It is hoped that the contributions made in the thesis research will make ANN appealing to more engineering or business applications in the future.

APPENDIX A: USER'S MANUAL FOR PINN TRAINER

Probabilistic Inference Neural Network (PINN) trainer is a generic neural network training and testing program developed based on a new NN scheme as proposed in Chapter 3 of the thesis.

Step 1. Prepare data and import data into trainer

The last column in a data table must be named as "Status", which flags the training/testing status for each record. Status 1 stands for a training record, and Status 2 for a testing record, and Status 0 for an ignored record. The next-to-last column in a data table must be named as "Output", storing Actual Output Values of the target risky variables such as actual production rates. All the remaining columns in a data table will be the input factors and no requirements are imposed on the names and relative order of columns. The trainer will automatically count the number of total inputs and read in data.

The prepared data table for PINN must be imported to the database file "PINN.mdb".

Step 2. Give a unique identifier key for a new training-testing trial

A unique identifier key is used to distinguish each training-testing scenario or trial, which is defined by the data source table to use, training / testing records within the data, the setup of the output zones, training parameters, and number of training iterations. Naming convention requires no space in the key and numbers and short texts are allowed such as "081899aWeld".

For a previous trial, user can pick out the identifier key from the drop-down list. Next, user may click the buttons on the switchboard to check training results

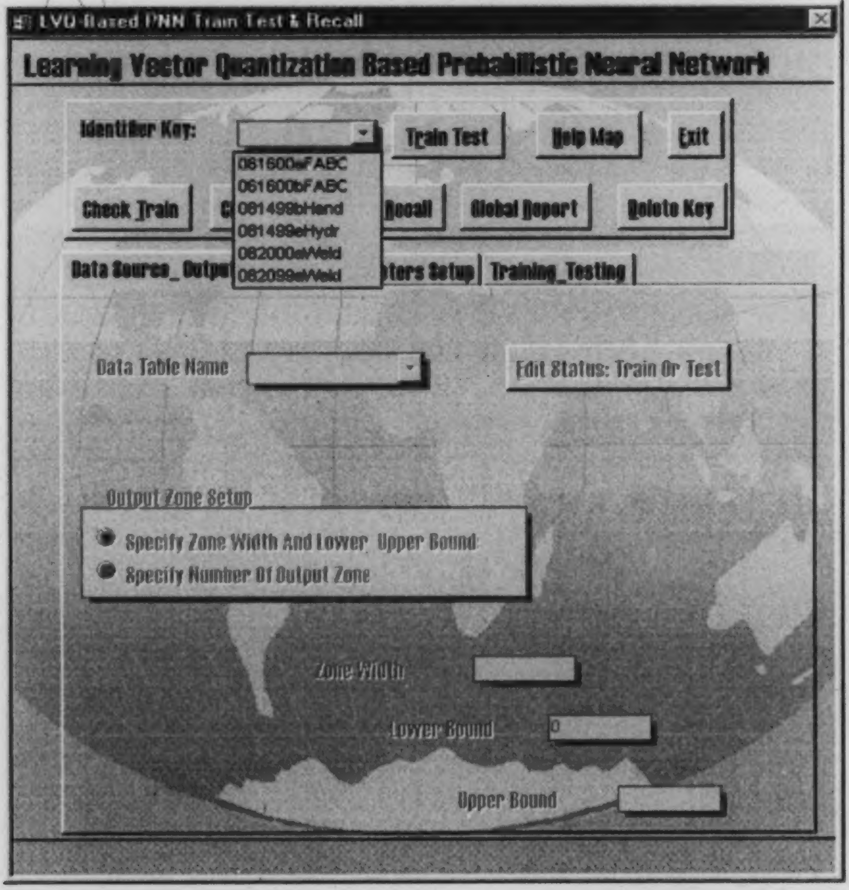
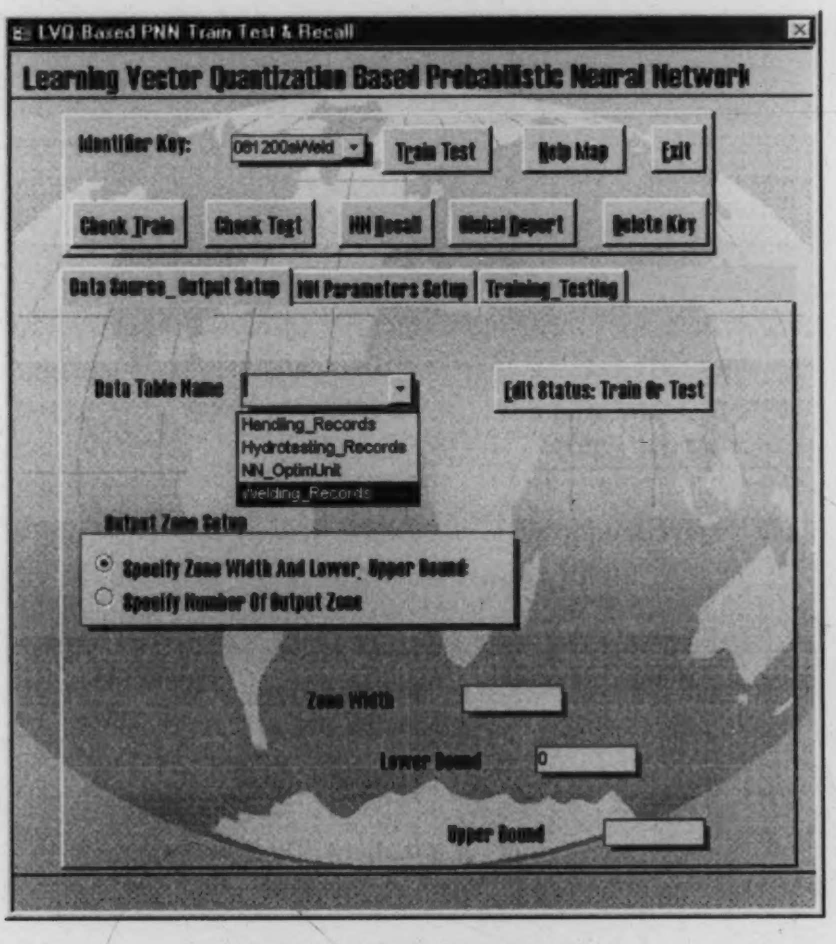

LVQ-Based PNN Train Test & Recall
Learning Vector Quantization Based Probabilistic Neural Network
Identifier Key:
061600aFABC
061600bFABC
081499bHand
081499eHydr
082000aWeld
082099aWeld
Train Test
Help Map
Exit
Check Train
Recall
Global Report
Delete Key
Data Source_ Output
ters Setup
Training_Testing
Data Table Name
Edit Status: Train Or Test
Output Zone Setup
● Specify Zone Width And Lower, Upper Bound
● Specify Number Of Output Zone
Zone Width
Lower Bound
0
Upper Bound

Figure A-1: Select an identifier key of one previous trial

("Check_Train" button), check testing result ("Check_Test" button), and check global report about training and testing ("Global Report" button) as shown in Figure A-1.

For a new trial, user needs to fill in a new identifier key in the Identifier Key box first. And then click "Train_Test" button on the switchboard to activate the program.

Step 3 Select data table, edit training / testing status and setup output zones


LVQ Based PNN Train Test & Recall
Learning Vector Quantization Based Probabilistic Neural Network
Train Test
Help Map
Exit
Check Train
Check Test
NN Recall
Global Report
Delete Key
Data Source_ Output Setup
NN Parameters Setup
Training_Testing
Data Table Name
Handling_Records
Hydrotesting_Records
NN_OptimUnit
Welding_Records
Edit Status: Train Or Test
Output Zone Setup
Zone Width
Upper Bound

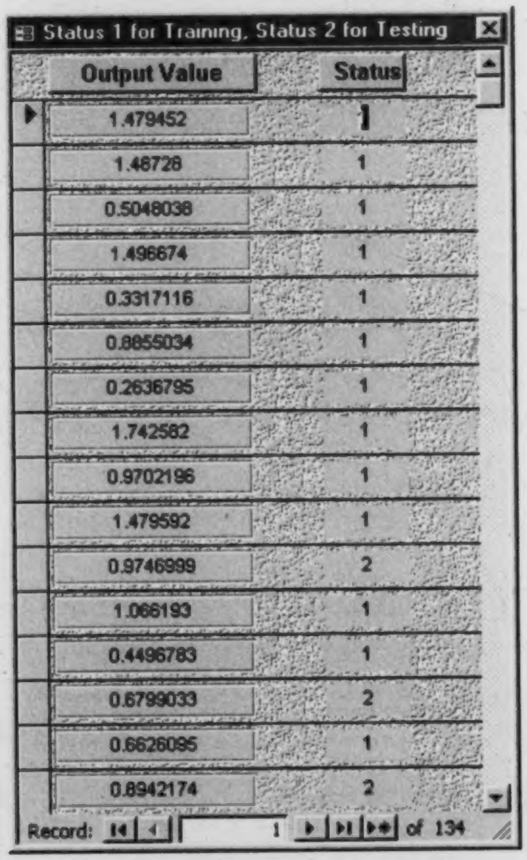
Identifier Key: 081200aWeld

● Specify Zone Width And Lower, Upper Bound
○ Specify Number Of Output Zone

Lower Bound: 0

Figure A-2: User selects data table

User selects one data table from the drop-down list of "Data Table Name" first (Figure A-2), and clicks the "Edit Status: Train or Test" button to flags training and testing records (Figure A-3). The trainer will read in data and display the maximum and minimum of the output values for user to setup the output zones.



Status 1 for Training, Status 2 for Testing

Output Value	Status
1.479452	1
1.46728	1
0.5048038	1
1.496674	1
0.3317116	1
0.8855034	1
0.2636795	1
1.742582	1
0.9702196	1
1.479592	1
0.9746999	2
1.066193	1
0.4496783	1
0.6799033	2
0.6626095	1
0.8942174	2

Record: 1 of 134

Figure A-3: Flag status of records

Setting up the output zones properly and adequately is crucial to PINN's performance. Too wide zone widths won't be adequate to help user make decision, while too narrow zone widths will probably sacrifice the accuracy of PINN's prediction. The following two issues should be taken into account:

- Precision requirement of user. Here is a heuristic formula to approximate zone width:

  Zone Width= 0.4*OutputRange*AccuracyThreshold

- Distribution of actual output values over the output zones. A uniform distribution generally yields better results.

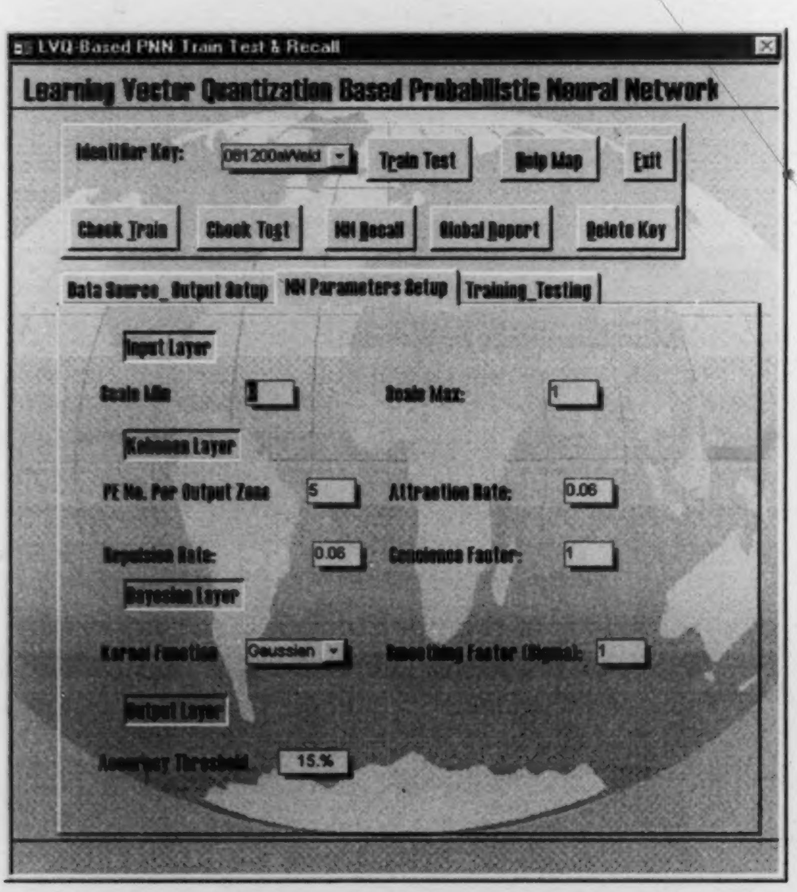
Two approaches are available to set up the output zones:

1. User specifies the number of output zones only. The trainer will evenly divide the actual output range into the number of output zones as user has specified, and automatically determine lower bound, upper bound, and mid value for each output zone.

2. User specifies the lower and upper bounds of the output range and the width of each zone as well. The trainer will start from the lower bound of output range and determine the boundaries of each output zone, until the upper bound of output value range is exceeded.

Step 4. Specify structure and learning parameters for PINN

Following the setup of output zones, user shifts focus to the next page to specify a number of structure and learning parameters for PINN including the scale max and min, the number of processing elements per output zone, the attraction rate and repulsion rate and conscience factor for learning, the smoothing factor for kernel function, and the accuracy threshold for performance measurement. Figure A-4 shows the program screen, user may take the default values or set new values for those parameters. Refer to the technical paper and online help for detailed explanations of


LVQ-Based PNN Train Test & Recall
Learning Vector Quantization Based Probabilistic Neural Network
Train Test
Help Map
Exit
Check Train
Check Test
NN Recall
Global Report
Delete Key
Data Source_ Output Setup
NN Parameters Setup
Training_Testing
Input Layer
Kohonen Layer
Bayesian Layer
Output Layer

Identifier Key: 081200aWeld

Scale Min: 0

Scale Max: 1

PE No. Per Output Zone: 5

Attraction Rate: 0.06

Repulsion Rate: 0.06

Concience Factor: 1

Kernel Function: Gaussian

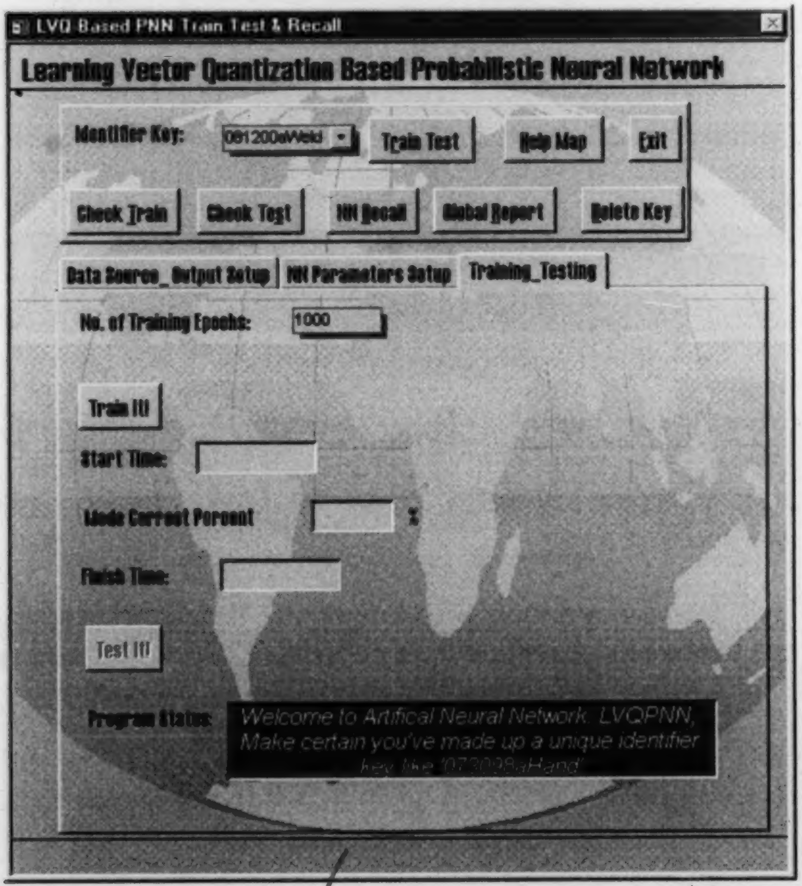
Smoothing Factor (Sigma): 1

Accuracy Threshold: 15.%

Figure A-4: Setup structure and learning parameters for PINN those parameters.

Step 5 Specify training iterations and train-test PINN

Following setting the PINN parameters, user shifts to the next page to specify


LVQ Based PNN Train Test & Recall
Learning Vector Quantization Based Probabilistic Neural Network
Train Test
Help Map
Exit
Check Train
Check Test
NN Recall
Global Report
Delete Key
Data Source_Output Setup
NN Parameters Setup
Training_Testing
Train It!
Start Time:
Mode Correct Percent
%
Finish Time:
Test It!

Identifier Key: 091200aWeld

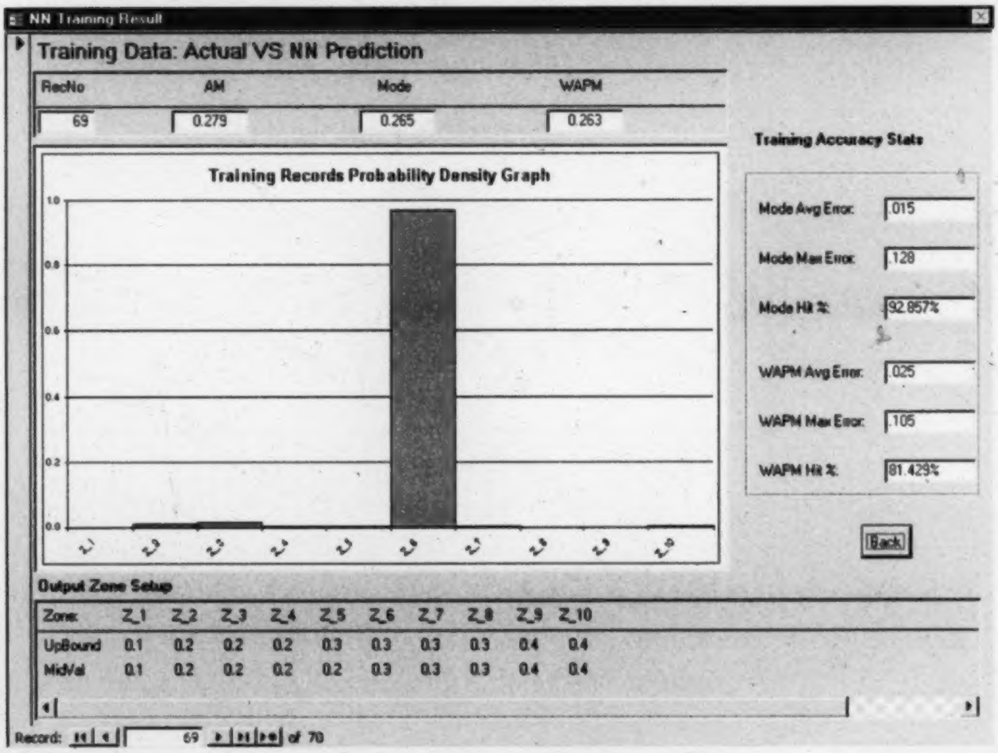
No. of Training Epochs: 1000

Program Status: Welcome to Artifical Neural Network. LVQPNN, Make certain you've made up a unique identifier key like '073098aHand'

Figure A-5: Specify training iterations and train-test PINN the training iterations by entering the "No. of Train Epochs", as shown in Figure A-5.

Step 6 Investigate whether the PINN model has been successfully trained

Following training and testing, user clicks the "Check_Train" button and the "Check_Test" button on the switch board to view the results for training data and


NN Training Result
Training Records Probability Density Graph
1.0
0.8
0.6
0.4
0.2
0.0
Back

Training Data: Actual VS NN Prediction

RecNo	AM	Mode	WAPM
69	0.279	0.265	0.263

Training Accuracy Stats

Mode Avg Error: .015
Mode Max Error: .128
Mode Hit %: 92.857%
WAPM Avg Error: .025
WAPM Max Error: .105
WAPM Hit %: 81.429%

Output Zone Setup

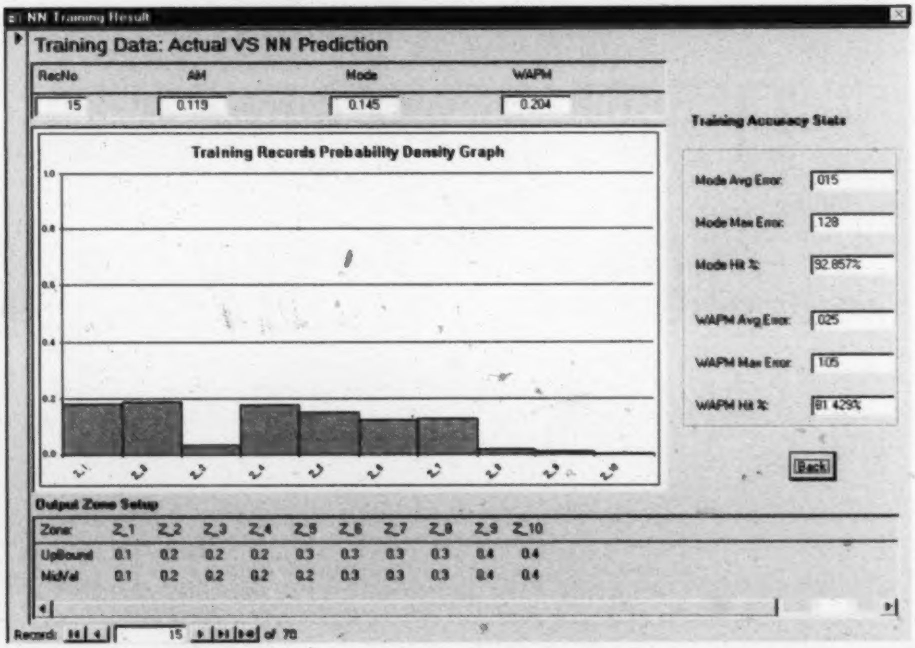
Zone:	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_10
UpBound	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.4
MidVal	0.1	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.4	0.4

Record: 69 of 70

Figure A-6: Check training results

testing data respectively and investigate whether the neural network has been successfully trained.

The majority of training records should indicate a centralized trend and the distribution generated by PINN falls in the correct output zones for their actual output values, as shown in Figure A-6. Otherwise, user should repeat from step 2 and perform another trial using different network structure and learning parameters, or increase learning iterations. Note that on the right side of Figure A-6, the accuracy statistics of point predictions are also included for user to judge the model' performance or maturity. User observes the testing data in a very similar manner. Testing data speaks more in


NN Training Result
Training Data: Actual VS NN Prediction
Training Records Probability Density Graph
1.0
0.8
0.6
0.4
0.2
0.0
Training Accuracy Stats
Back
Output Zone Setup
Record:
of 70

RecNo: 15
AM: 0.119
Mode: 0.145
WAPM: 0.204

Mode Avg Error: 015
Mode Max Error: 128
Mode Hit %: 92.857%
WAPM Avg Error: 025
WAPM Max Error: 105
WAPM Hit %: 81.429%

Zone:	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_10
UpBound	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.4
MidVal	0.1	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.4	0.4

Record: 15 of 70

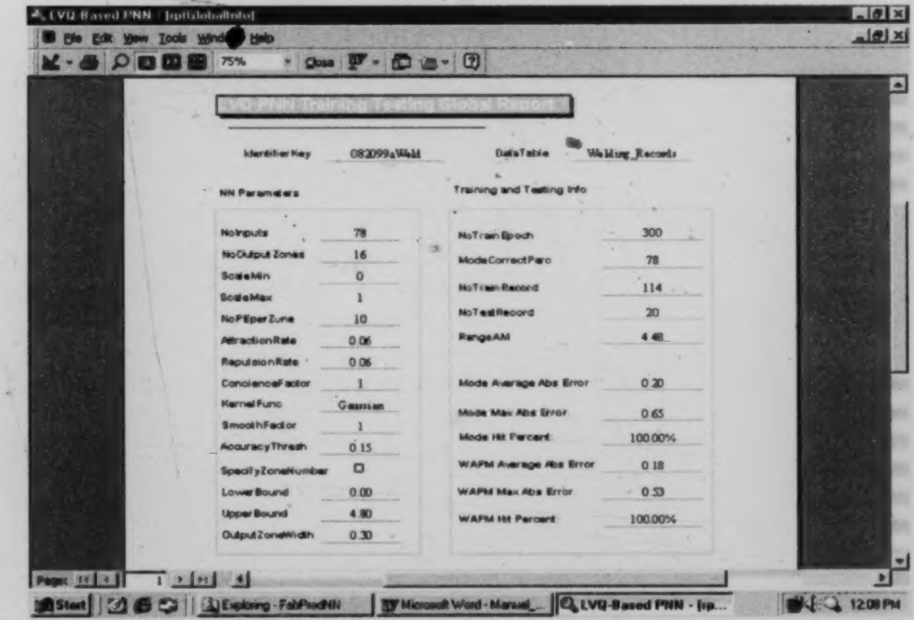
Figure A-7: Detected noise in training data

judging the network's performance than training data because the trainer has not seen the testing records in the learning process.

In case that after a number of different train-test trials, for a particular training record, the PINN indicates a very far-off point prediction value (mode) comparing with the actual output value, or the PINN demonstrates a very dispersed distribution as shown in Figure A-7, such a record is likely to be noise in the data and the data of the record should be examined for errors.

Step 7 Recall based on a trained PINN model

Once satisfactory results are obtained for both training and testing data, the


LVQ Based PNN - [rptGlobalInfo]
File Edit View Tools Window Help
75%
Close
Page: 1
Start
Exploring - FabPredNN
Microsoft Word - Manual_...
LVQ-Based PNN - [rp...
12:08 PM

LVQ PNN Training Testing Global Report

IdentifierKey: 082099aWeld
DataTable: Welding_Records

NN Parameters

NoInputs: 78
NoOutputZones: 16
ScaleMin: 0
ScaleMax: 1
NoPEperZone: 10
AttractionRate: 0.06
RepulsionRate: 0.06
ConcienceFactor: 1
KernelFunc: Gaussian
SmoothFactor: 1
AccuracyThresh: 0.15
SpecifyZoneNumber: ☐
LowerBound: 0.00
UpperBound: 4.80
OutputZoneWidth: 0.30

Training and Testing Info

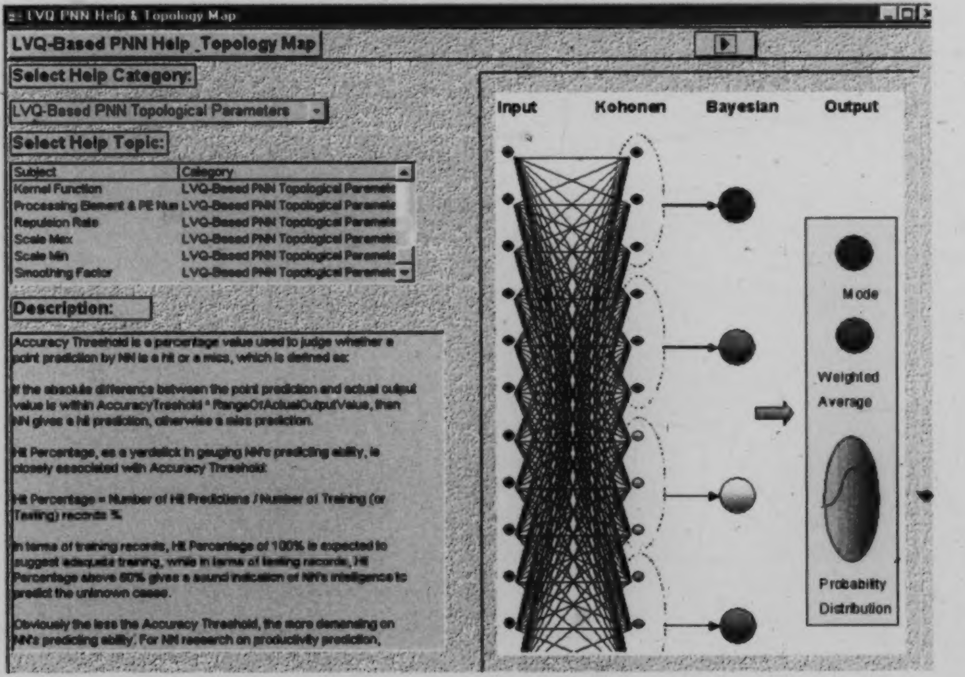
NoTrainEpoch: 300
ModeCorrectPerc: 78
NoTrainRecord: 114
NoTestRecord: 20
RangeAM: 4.48
Mode Average Abs Error: 0.20
Mode Max Abs Error: 0.65
Mode Hit Percent: 100.00%
WAPM Average Abs Error: 0.18
WAPM Max Abs Error: 0.53
WAPM Hit Percent: 100.00%

Figure A-8: Global report for a train-test trial

PINN model is declared to be trained and ready for developing a recall program. User clicks the "Global Report" button to review the information about this trial as shown in Figure A-8.

If user does not want to keep a trial any more, select the identifier key for the trial and click the "Delete Key" button on the switchboard to delete all the data related to the trial.

An on-line help is also developed giving details about how to use the trainer


LVQ PNN Help & Topology Map
LVQ-Based PNN Help _Topology Map
Select Help Category:
LVQ-Based PNN Topological Parameters
Select Help Topic:
Description:
Input
Kohonen
Bayesian
Output
Mode
Weighted Average
Probability Distribution

Subject	Category
Kernel Function	LVQ-Based PNN Topological Paramete
Processing Element & PE Nu	LVQ-Based PNN Topological Paramete
Repulsion Rate	LVQ-Based PNN Topological Paramete
Scale Max	LVQ-Based PNN Topological Paramete
Scale Min	LVQ-Based PNN Topological Paramete
Smoothing Factor	LVQ-Based PNN Topological Paramete

Accuracy Threshold is a percentage value used to judge whether a point prediction by NN is a hit or a miss, which is defined as:

If the absolute difference between the point prediction and actual output value is within AccuracyTreshold * RangeOfActualOutputValue, then NN gives a hit prediction, otherwise a miss prediction.

Hit Percentage, as a yardstick in gauging NN's predicting ability, is closely associated with Accuracy Threshold:

Hit Percentage = Number of Hit Predictions / Number of Training (or Testing) records %

In terms of training records, Hit Percentage of 100% is expected to suggest adequate training, while in terms of testing records, Hit Percentage above 80% gives a sound indication of NN's intelligence to predict the unknown cases.

Obviously the less the Accuracy Threshold, the more demanding on NN's predicting ability. For NN research on productivity prediction,

Figure A-9: PINN Trainer on-line help

program along with some technical descriptions about the PINN model, as shown in

Figure A-9. By selecting a category, related topics are filtered out; user chooses one topic of interest, the description will be automatically displayed.

APPENDIX B: USERS' MANUAL FOR FABMASTER

FabMaster is a historical project data warehousing system customized for the Fabrication Facilities of PCL Industrial Constructors, Inc. It is an automated data processing tool to extract raw data from Fabrication Resources Planning System, Weld Tracking System, Labor Cost Control System, and convert raw data into aggregate quantity data at spool level and various ratios of productivity, quality control and configuration complexity at cost-center level. A cost center, defined by project number, material type and size range of spools, is the level of detail that actual labor hours were tracked in the corporate labor cost control system. Two levels of compilation are involved in FabMaster to convert raw data into item-coded productivity information, i.e. the spool level and the cost-center level. The coding systems for material type and size range of spools used in the company and FabMaster are shown in Table B-1 and B-2. The item codes for spool level compilation are shown in Table B-3.

Table B-1: Size Range Codes

ID	Description
1	< 2"
2	2-4"
6	6-14"
16	16-24"
30	30-48"
0	Total

Table B-2: Material Type Codes

ID	Description
1	CS (Carbon)
2	SS (Stainless)
3	AL (Aluminum)
4	AS (Alloy)
0	Total

Table B-3: Item Codes for Spool Level Data Compilation

ID	Item_Code	Description
1	11	No. of pipe pieces
2	12	Footage of pipe pieces
3	13	Diameter Inch Ft of pipe pieces
4	14	Tons of pipe pieces
5	21	No. of pipe pieces longer than 6 ft
6	22	Footage of pipe pieces longer than 6 ft
7	23	DiaInFt of pipe pieces longer than 6 ft
8	24	Tons of pipe pieces longer than 6 ft
9	31	No. of Stub Ends
10	41	No. of Branches (Olets)
11	51	No. of Caps/Plugs
12	61	No. of Elbows
13	71	No. of Swages
14	81	No. of Blind Flanges
15	91	No. of Dummy Legs
16	101	No. of SW/TH Couplings
17	111	No. of Lap Joint Flanges
18	121	No. of Anchors/Shoes/Slider Supports
19	131	No. of Nipples
20	141	No. of Orifice Flange
21	151	No. of Reducers
22	161	No. of Slip-on/SW/TH Flange
23	171	No. of Tees
24	181	No. of Unions
25	191	No. of Valves
26	201	No. of Weld Neck Flanges
27	211	No. of Laterals
28	221	No. of Misc Items
29	231	No. of Flanges
30	241	No. of In-Line Fittings
31	251	No. of Out-Line Fittings
32	261	No. of Supports
33	271	No. of Design Welds
34	272	Diameter Inch of Design Welds
35	273	Equivalent Diameter Inch of Design Welds

ID	Item_Code	Description
36	274	Volume of Design Welds
37	281	No. of BW Design Welds
38	282	Diameter Inch of BW Design Welds
39	283	Equivalent Diameter Inch of BW Design Welds
40	284	Volume of BW Design Welds
41	291	No. of SW Design Welds
42	292	Diameter Inch of SW Design Welds
43	293	Equivalent Diameter Inch of SW Design Welds
44	294	Volume of SW Design Welds
45	301	No. of OL Design Welds
46	302	Diameter Inch of OL Design Welds
47	303	Equivalent Diameter Inch of OL Design Welds
48	304	Volume of OL Design Welds
49	311	No. of Pressure Attachments in Design Welds
50	312	Diameter Inch of Pressure Attachments in Design Welds
51	313	Equivalent Diameter Inch of Pressure Attachments in Design Welds
52	314	Volume of Pressure Attachments in Design Welds
53	321	No. of Non Pressure Attachments in Design Welds
54	322	Diameter Inch of Non Pressure Attachments in Design Welds
55	323	Equivalent Diameter Inch of No Pressure Attachments in Design Welds
56	324	Volume of No Pressure Attachments in Design Welds
57	331	No. of Positon Welds in Design Welds
58	332	Diameter Inch of Positon Welds in Design Welds
59	333	Equivalent Diameter Inch of Positon Welds in Design Welds
60	334	Volume of Positon Welds in Design Welds
61	341	No. of Multi-station Roll Welds in Design Welds
62	342	Diameter Inch of Multi-station Roll Welds in Design Welds
63	343	Equivalent Diameter Inch of Multi-station Roll Welds in Design Welds
64	344	Volume of Multi-station Roll Welds in Design Welds
65	354	Volume of Tig Process in Design welds
66	364	Volume of Mig Process in Design Welds
67	374	Volume of FCAW Process in Design Welds
68	384	Volume of Stick Process in Design Welds
69	394	Volume of SubArc Process in Design Welds
70	404	Volume of Rotoweld Process in Design Welds
71	411	No. of Reworked Welds
72	412	Diameter Inch of Reworked Welds
73	413	Equivalent Diameter Inch of Reworked Welds

ID	Item_Code	Description
74	414	Volume of Reworked Welds
75	422	No. of Cut Sheet Revisions
76	434	Spool Weight in Tons
77	441	RT percent per Spec
78	442	MT percent per Spec
79	443	PT percent per Spec
80	444	PMI percent per Spec
81	445	PWHT 1/0 per Spec
82	446	FT percent per Spec
83	447	BHN
84	448	VT percent per Spec
85	449	UT percent per Spec
86	451	No. of Accepted Welds
88	435	Weld Units in Spool
89	464	Weight of Non-pipe in Spool
90	1	No of Spools in a material-size group

Step 1 Download Raw Data from Corporate Management Systems into FabMaster

Six raw data tables of a project required for FabMaster to process are directly downloaded from the corporate databases of various management systems in electronic formats, namely RD_BranchPlant_SP (Spool) and RD_BranchPlant_MT (Pieces) from J. D. Edwards material resources planning system, plus RD_BranchPlant_DG (Drawing), RD_BranchPlant_SC (Spec), RD_BranchPlant_WD (Weld Details) and RD_BranchPlant_WW (Weld Welders) from WeldTrack quality control system, as shown in the program flow chart of FabMaster (Figure B-1).

User enters the project number and clicks the "Link RD Tables" to import raw data tables. If all needed tables are in place, user kicks off the program flow by hitting the "Process it" button.

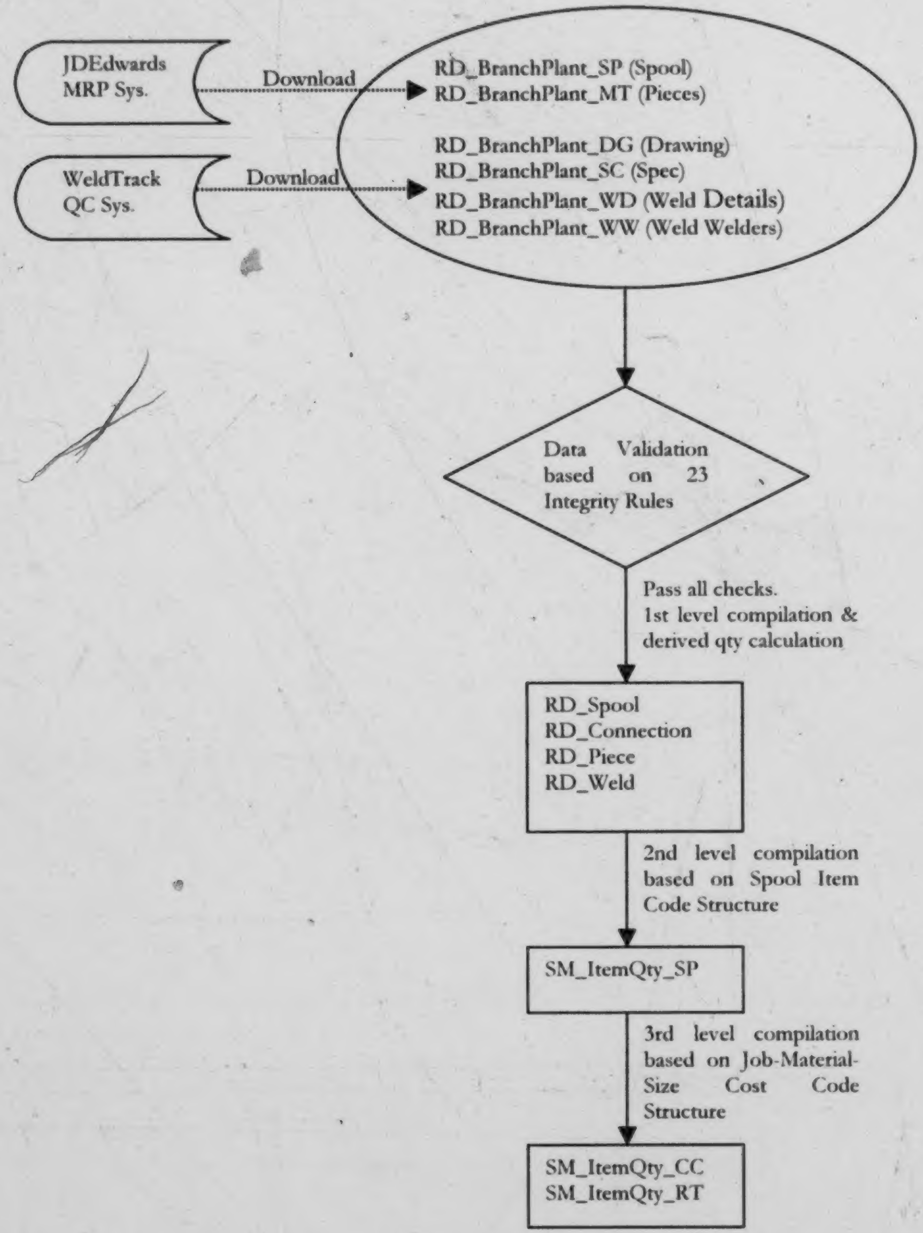

JDEdwards
MRP Sys.
Download
WeldTrack
QC Sys.
Download
RD_BranchPlant_SP (Spool)
RD_BranchPlant_MT (Pieces)
RD_BranchPlant_DG (Drawing)
RD_BranchPlant_SC (Spec)
RD_BranchPlant_WD (Weld Details)
RD_BranchPlant_WW (Weld Welders)
Data Validation based on 23 Integrity Rules
Pass all checks.
1st level compilation & derived qty calculation
RD_Spool
RD_Connection
RD_Piece
RD_Weld
2nd level compilation based on Spool Item Code Structure
SM_ItemQty_SP
3rd level compilation based on Job-Material-Size Cost Code Structure
SM_ItemQty_CC
SM_ItemQty_RT

Figure B-1: Program Flow Chart of FabMaster

Step 2. Data Validation Based on Pre-defined Rules

The following rules were programmed in FabMaster to detect abnormalities and prompt user to scrub raw data of errors prior to processing.

1. No blank is allowed in [Item Number] in RD_BranchPlant_MT.
2. No blank is allowed in [Weld Process] in RD_BranchPlant_WD.
3. [Weld Process] in RD_BranchPlant_WD must be a known process-combination in table LP_WeldProcessCombo.
4. No blank is allowed in [Joint Type] in RD_BranchPlant_WD.
5. [Joint Type] in RD_BranchPlant_WD must be a known type in table LP_JointType_PANP.
6. No blank or 0 is allowed in [Weld Size] in RD_BranchPlant_WD.
7. [Weld Size] in RD_BranchPlant_WD must be a known size in table LP_PipeOD.
8. No blank or 0 is allowed in [Weld Thickness] in RD_BranchPlant_WD.
9. No redundant spool is allowed to exist in RD_BranchPlant_SP.
10. No blank or 0 is allowed in [Spool Weight] in RD_BranchPlant_SP.
11. No blank or 0 is allowed in [Spool Units] in RD_BranchPlant_SP.
12. No blank or 0 is allowed in [Size Group] in RD_BranchPlant_SP.

13. [Size Group] in RD_BranchPlant_SP must be a known one in table LP_SizeGroup.

14. No blank or 0 is allowed in [Material Group] in RD_BranchPlant_SP.

15. [Material Group] in RD_BranchPlant_SP must be a known one in table LP_MaterialGroup.

16. [Spool Number] in RD_BranchPlant_SP must have a corresponding record in RD_BranchPlant_DG.

17. No blank is allowed in [Spec] in RD_BranchPlant_SC.

18. [Spec] in RD_BranchPlant_DG must have a corresponding reference in RD_BranchPlant_SC.

19. [Spool Number] in RD_BranchPlant_SP must have a corresponding record in RD_BranchPlant_MT.

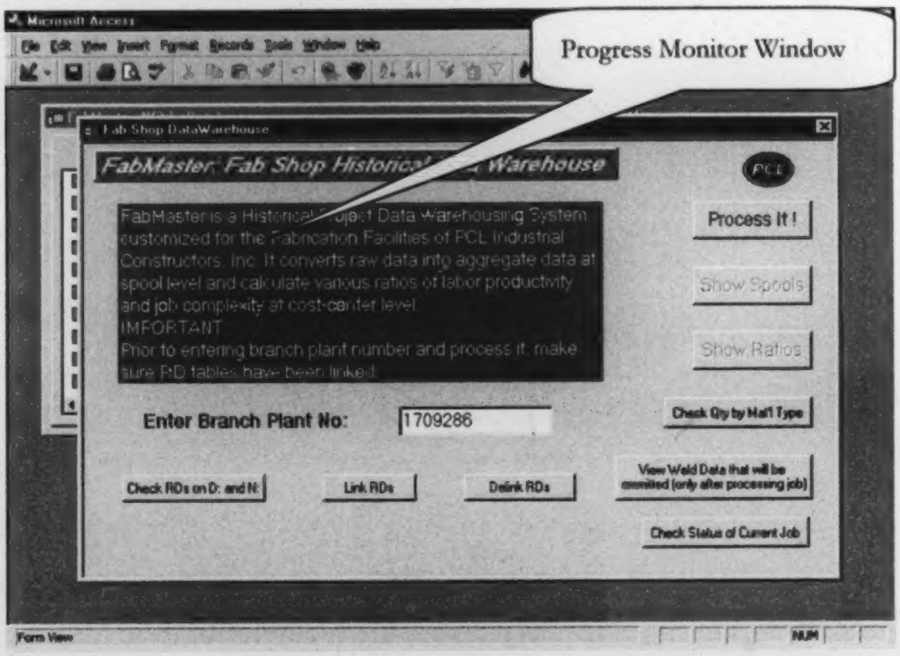
20. [Weld Thichness] in RD_BranchPlant_WD like 3000/6000 must be able to be converted to inches by finding a reference in LP_PipeThickness.

21. [Weld Thichness] in RD_BranchPlant_WD must not be abnormally large (greater than 5").

22. [Weld Size] in RD_BranchPlant_WD must be able to be converted to Equivalent Diameter Inches by finding a reference in LP_PipeThickness based on size and thickness.

23. For an Olet type weld, a reference in LP_OletDim based on weld size must be found.

L


Microsoft Access
File Edit View Insert Format Records Tools Window Help
Progress Monitor Window
Fab Shop DataWarehouse
FabMaster: Fab Shop Historical
Warehouse
PCL
Process It !
Show Spools
Show Ratios
Check Qty by Mat'l Type
View Weld Data that will be ommited (only after processing job)
Check RDs on D: and N:
Link RDs
Delink RDs
Check Status of Current Job
Form View
NUM

FabMaster is a Historical roject Data Warehousing System customized for the abrication Facilities of PCL Industrial Constructors, Inc. It converts raw data into aggregate data at spool level and calculate various ratios of labor productivity and job complexity at cost-center level.
IMPORTANT
Prior to entering branch plant number and process it, make sure RD tables have been linked.

Enter Branch Plant No: 1709286

Figure B-2: Main User Interface of FabMaster

FabMaster will hint user about the detected problem records, violated rules and solutions (either update cross-reference tables or correct raw data tables) in its progress monitor windows as shown in Figure B-2. To resume the program flow after fixing problems, user needs to hit the “process it” button again to continue the process from where it paused last time.

Step 3 Unitize a project and perform spool level compilation

Once all the checks on raw data are passed, FabMaster runs its built-in programs to automatically unitize a project into “Fabrication units”, compute the quantities of various work items in pre-specified units of measure, and store the results into four temporary tables, namely RD_Spool, RD_Piece, RD_Connection, and RD_Weld. The first level compilation is conducted based on those temporary tables to generate the item-coded aggregate data for each spool and appended to spool summary table "SM_ItemQty_SP".

Step 4. Compile data at cost center level and compute ratios

The cost-center level data compilation and ratio computation follows the spool level data processing and the results will be appended to two summary tables "SM_ItemQty_CC" for aggregate quantities, and "SM_ItemQty_RT" for final ratios. Table B-4 shows samples of SM_ItemQty_RT" based on one small project with one material type and one size range only. A big project often has more than one material types and size ranges of spools. FabMaster will generate valid ratios only for a specific project and automatically handle the roll-up of ratios to various total levels.

Table B-4: Sample of FabMaster Outputs

Material	Size	Ratio Description	Ratio
Total	Total	Total ManHours / DiaIn*Ft	1.705658E-04
Total	Total	Total ManHours / Equiv.DiaIn*Ft	1.705658E-04
Total	Total	Total ManHours / Equiv.DiaIn	0.6160267
Total	Total	Total ManHours / Volume	2.657094
Total	Total	Total ManHours / Unit	0.1994056
Total	Total	No. of Pipe Pieces / Footage	5.316105E-02
Total	Total	No. of Pipe Pieces / DiaInFt	8.862623E-03
Total	Total	No. of Pipe Pieces / Ton	5.599411
Total	Total	No. of Pipe Pieces / Unit	5.187787E-02
Total	Total	No. of Pipe Pieces Over 3 ft / Footage	5.066913E-02
Total	Total	No. of Pipe Pieces Over 3 ft / DiaInFt	8.447187E-03
Total	Total	No. of Pipe Pieces Over 3 ft / Ton	5.336939
Total	Total	No. of Pipe Pieces over 3 ft / Unit	4.944609E-02
Total	Total	No. of Flanges / Footage	5.53761E-03
Total	Total	No. of Flanges / DiaInFt	9.231899E-04
Total	Total	No. of Flanges / Ton	0.583272
Total	Total	No. of Flanges / Unit	5.403945E-03
Total	Total	No. of In-Line Fittings / Footage	1.107522E-03
Total	Total	No. of In-Line Fittings / DiaInFt	1.84638E-04
Total	Total	No. of In-Line Fittings / Ton	0.1166544
Total	Total	No. of In-Line Filtings / Unit	1.080789E-03
Total	Total	No. of Non-In-Line Fittings / Footage	5.205353E-02
Total	Total	No. of Non-In-Line Fittings / DiaInFt	8.677985E-03
Total	Total	No. of Non-In-Line Fittings / Ton	5.482757
Total	Total	No. of Non-In-Line Filtings / Unit	5.079708E-02
Total	Total	No. of Valves / Footage	0
Total	Total	No. of Valves / DiaInFt	0
Total	Total	No. of Valves / Ton	0
Total	Total	No.of Valves / Unit	0
Total	Total	No. of Supports / Footage	0
Total	Total	No. of Supports / DiaInFt	0
Total	Total	No. of Supports / Ton	0
Total	Total	No. of Supports / Unit	0
Total	Total	No. of Misc. / Footage	5.53761E-04

Material	Size	Ratio Description	Ratio
Total	Total	No. of Misc. / DiaInFt	9.231899E-05
Total	Total	No. of Misc. / Ton	0.0583272
Total	Total	No.of Misc. / Unit	5.403945E-04
Total	Total	Pipe Weight / Spool Weight	0.9404383
Total	Total	Non-Pipe Weight / Spool Weight	5.956174E-02
Total	Total	No. of Connections (Design Welds) / Footage	5.786802E-02
Total	Total	No. of Connections (Design Welds) / DiaInFt	9.647334E-03
Total	Total	No. of Connections (Design Welds) / Ton	6.095192
Total	Total	No. of Connections (Design Welds) / Unit	5.647122E-02
Total	Total	No. of Multi-Station Roll Welds / Footage	2.353484E-02
Total	Total	No. of Multi-Station Roll Welds / DiaInFt	3.923557E-03
Total	Total	No. of Multi-Station Roll Welds / Ton	2.478906
Total	Total	No. of Multi-Station Roll Welds / Unit	2.296676E-02
Total	Total	No. of Repaired Welds / Footage	3.322566E-03
Total	Total	No. of Repaired Welds / DiaInFt	5.539139E-04
Total	Total	No. of Repaired Welds / Ton	0.3499632
Total	Total	No. of Repair Welds / Unit	3.242367E-03
Total	Total	BW DiaIn / Design Weld DiaIn	0.9782972
Total	Total	BW Equiv.DiaIn / Design Weld Equiv.DiaIn	0.9782972
Total	Total	BW Vol. / Design Weld Vol.	0.9940413
Total	Total	SW DiaIn / Design Weld DiaIn	1.335559E-02
Total	Total	SW Equiv.DiaIn / Design Weld Equiv.DiaIn	1.335559E-02
Total	Total	SW Vol. / Design Weld Vol.	3.170402E-03
Total	Total	OL DiaIn / Design Weld DiaIn	8.347246E-03
Total	Total	OL Equiv.DiaIn / Design Weld Equiv.DiaIn	8.347246E-03
Total	Total	OL Vol. / Design Weld Vol.	2.788416E-03
Total	Total	Pressure Attachment / Design Weld DiaIn	0.9866444
Total	Total	Pressure Attachment / Design Weld Equiv.DiaIn	0.9866444
Total	Total	Pressure Attachment / Design Weld Vol.	0.9968296
Total	Total	Non Pressure Attachment / Design Weld DiaIn	1.335559E-02
Total	Total	Non Pressure Attachment / Design Weld Equiv.DiaIn	1.335559E-02
Total	Total	Non Pressure Attachment / Design Weld Vol.	3.170402E-03
Total	Total	Position Weld / Design Weld DiaIn	0
Total	Total	Position Weld / Design Weld Equiv.DiaIn	0
Total	Total	Position Weld / Design Weld Vol.	0
Total	Total	Roll Weld / Design Weld DiaIn	1
Total	Total	Roll Weld / Design Weld Equiv.DiaIn	1
Total	Total	Roll Weld / Design Weld Vol.	1
Total	Total	Multi-Station Roll Weld / Design Weld DiaIn	0.4257095

Material	Size	Ratio Description	Ratio
Total	Total	Multi-Station Roll Weld / Design Weld Equiv.DiaIn	0.4257095
Total	Total	Multi-Station Roll Weld / Design Weld Vol.	0.432653
Total	Total	Single-Station Roll Weld / Design Weld DiaIn	0.5742905
Total	Total	Single-Station Roll Weld / Design Weld Equiv.DiaIn	0.5742905
Total	Total	Single-Station Roll Weld / Design Weld Vol.	0.567347
Total	Total	Tig Process Weld / Design Weld Vol.	0.1067211
Total	Total	Mig Process Weld / Design Weld Vol.	0
Total	Total	FCAW Process Weld / Design Weld Vol.	0
Total	Total	Stick Process Weld / Design Weld Vol.	0.893279
Total	Total	SubArc Process Weld / Design Weld Vol.	0
Total	Total	Rotoweld Process Weld / Design Weld Vol.	0
Total	Total	Repair Rate (No. of R / R+A)	2.933985E-02
Total	Total	No. of Cut Sheet Revision / No. of Spool	0
Total	Total	RT rate /Spool	100
Total	Total	MT rate /Spool	100
Total	Total	PT rate /Spool	0
Total	Total	PMI rate /Spool	100
Total	Total	PWHT rate /Spool	1
Total	Total	FT rate /Spool	0
Total	Total	BHN rate /Spool	100
Total	Total	VT rate /Spool	100
Total	Total	UT rate /Spool	100
Total	Total	Non-Welded Spool/Welded Spool (Weight)	0
Total	Total	Non-Welded Spool/Welded Spool (Units)	0
Total	6-14"	No. of Pipe Pieces / Footage	5.316105E-02
Total	6-14"	No. of Pipe Pieces / DiaInFt	8.862623E-03
Total	6-14"	No. of Pipe Pieces / Ton	5.599411
Total	6-14"	No. of Pipe Pieces / Unit	5.187787E-02
Total	6-14"	No. of Pipe Pieces Over 3 ft / Footage	5.066913E-02
Total	6-14"	No. of Pipe Pieces Over 3 ft / DiaInFt	8.447187E-03
Total	6-14"	No. of Pipe Pieces Over 3 ft / Ton	5.336939
Total	6-14"	No. of Pipe Pieces over 3 ft / Unit	4.944609E-02
Total	6-14"	No. of Flanges / Footage	5.53761E-03
Total	6-14"	No. of Flanges / DiaInFt	9.231899E-04
Total	6-14"	No. of Flanges / Ton	0.583272
Total	6-14"	No. of Flanges / Unit	5.403945E-03
Total	6-14"	No. of In-Line Fittings / Footage	1.107522E-03
Total	6-14"	No. of In-Line Fittings / DiaInFt	1.84638E-04
Total	6-14"	No. of In-Line Fittings / Ton	0.1166544

Material	Size	Ratio Description	Ratio
Total	6-14"	No. of In-Line Filtings / Unit	1.080789E-03
Total	6-14"	No. of Non-In-Line Fittings / Footage	5.205353E-02
Total	6-14"	No. of Non-In-Line Fittings / DiaInFt	8.677985E-03
Total	6-14"	No. of Non-In-Line Fittings / Ton	5.482757
Total	6-14"	No. of Non-In-Line Filtings / Unit	5.079708E-02
Total	6-14"	No. of Valves / Footage	0
Total	6-14"	No. of Valves / DiaInFt	0
Total	6-14"	No. of Valves / Ton	0
Total	6-14"	No.of Valves / Unit	0
Total	6-14"	No. of Supports / Footage	0
Total	6-14"	No. of Supports / DiaInFt	0
Total	6-14"	No. of Supports / Ton	0
Total	6-14"	No. of Supports / Unit	0
Total	6-14"	No. of Misc. / Footage	5.53761E-04
Total	6-14"	No. of Misc. / DiaInFt	9.231899E-05
Total	6-14"	No. of Misc. / Ton	0.0583272
Total	6-14"	No.of Misc. / Unit	5.403945E-04
Total	6-14"	Pipe Weight / Spool Weight	0.9404383
Total	6-14"	Non-Pipe Weight / Spool Weight	5.956174E-02
Total	6-14"	No. of Connections (Design Welds) / Footage	5.786802E-02
Total	6-14"	No. of Connections (Design Welds) / DiaInFt	9.647334E-03
Total	6-14"	No. of Connections (Design Welds) / Ton	6.095192
Total	6-14"	No. of Connections (Design Welds) / Unit	5.647122E-02
Total	6-14"	No. of Multi-Station Roll Welds / Footage	2.353484E-02
Total	6-14"	No. of Multi-Station Roll Welds / DiaInFt	3.923557E-03
Total	6-14"	No. of Multi-Station Roll Welds / Ton	2.478906
Total	6-14"	No. of Multi-Station Roll Welds / Unit	2.296676E-02
Total	6-14"	No. of Repaired Welds / Footage	3.322566E-03
Total	6-14"	No. of Repaired Welds / DiaInFt	5.539139E-04
Total	6-14"	No. of Repaired Welds / Ton	0.3499632
Total	6-14"	No. of Repair Welds / Unit	3.242367E-03
Total	6-14"	BW DiaIn / Design Weld DiaIn	0.9782972
Total	6-14"	BW Equiv.DiaIn / Design Weld Equiv.DiaIn	0.9782972
Total	6-14"	BW Vol. / Design Weld Vol.	0.9940413
Total	6-14"	SW DiaIn / Design Weld DiaIn	1.335559E-02
Total	6-14"	SW Equiv.DiaIn / Design Weld Equiv.DiaIn	1.335559E-02
Total	6-14"	SW Vol. / Design Weld Vol.	3.170402E-03
Total	6-14"	OL DiaIn / Design Weld DiaIn	8.347246E-03
Total	6-14"	OL Equiv.DiaIn / Design Weld Equiv.DiaIn	8.347246E-03

Material	Size	Ratio Description	Ratio
Total	6-14"	OL Vol. / Design Weld Vol.	2.788416E-03
Total	6-14"	Pressure Attachment / Design Weld DiaIn	0.9866444
Total	6-14"	Pressure Attachment / Design Weld Equiv.DiaIn	0.9866444
Total	6-14"	Pressure Attachment / Design Weld Vol.	0.9968296
Total	6-14"	Non Pressure Attachment / Design Weld DiaIn	1.335559E-02
Total	6-14"	Non Pressure Attachment / Design Weld Equiv.DiaIn	1.335559E-02
Total	6-14"	Non Pressure Attachment / Design Weld Vol.	3.170402E-03
Total	6-14"	Position Weld / Design Weld DiaIn	0
Total	6-14"	Position Weld / Design Weld Equiv.DiaIn	0
Total	6-14"	Position Weld / Design Weld Vol.	0
Total	6-14"	Roll Weld / Design Weld DiaIn	1
Total	6-14"	Roll Weld / Design Weld Equiv.DiaIn	1
Total	6-14"	Roll Weld / Design Weld Vol.	1
Total	6-14"	Multi-Station Roll Weld / Design Weld DiaIn	0.4257095
Total	6-14"	Multi-Station Roll Weld / Design Weld Equiv.DiaIn	0.4257095
Total	6-14"	Multi-Station Roll Weld / Design Weld Vol.	0.432653
Total	6-14"	Single-Station Roll Weld / Design Weld DiaIn	0.5742905
Total	6-14"	Single-Station Roll Weld / Design Weld Equiv.DiaIn	0.5742905
Total	6-14"	Single-Station Roll Weld / Design Weld Vol.	0.567347
Total	6-14"	Tig Process Weld / Design Weld Vol.	0.1067211
Total	6-14"	Mig Process Weld / Design Weld Vol.	0
Total	6-14"	FCAW Process Weld / Design Weld Vol.	0
Total	6-14"	Stick Process Weld / Design Weld Vol.	0.893279
Total	6-14"	SubArc Process Weld / Design Weld Vol.	0
Total	6-14"	Rotoweld Process Weld / Design Weld Vol.	0
Total	6-14"	Repair Rate (No. of R / R+A)	2.933985E-02
Total	6-14"	No. of Cut Sheet Revision / No. of Spool	0
Total	6-14"	RT rate /Spool	100
Total	6-14"	MT rate /Spool	100
Total	6-14"	PT rate /Spool	0
Total	6-14"	PMI rate /Spool	100
Total	6-14"	PWHT rate /Spool	1
Total	6-14"	FT rate /Spool	0
Total	6-14"	BHN rate /Spool	100
Total	6-14"	VT rate /Spool	100
Total	6-14"	UT rate /Spool	100
Total	6-14"	Non-Welded Spool/Welded Spool (Weight)	0
Total	6-14"	Non-Welded Spool/Welded Spool (Units)	0
AS (Alloy)	Total	Total ManHours / DiaIn*Ft	1.705658E-04

Material	Size	Ratio Description	Ratio
AS (Alloy)	Total	Total ManHours / Equiv.DiaIn*Ft	1.705658E-04
AS (Alloy)	Total	Total ManHours / Equiv.DiaIn	0.6160267
AS (Alloy)	Total	Total ManHours / Volume	2.657094
AS (Alloy)	Total	Total ManHours / Unit	0.1994056
AS (Alloy)	Total	No. of Pipe Pieces / Footage	5.316105E-02
AS (Alloy)	Total	No. of Pipe Pieces / DiaInFt	8.862623E-03
AS (Alloy)	Total	No. of Pipe Pieces / Ton	5.599411
AS (Alloy)	Total	No. of Pipe Pieces / Unit	5.187787E-02
AS (Alloy)	Total	No. of Pipe Pieces Over 3 ft / Footage	5.066913E-02
AS (Alloy)	Total	No. of Pipe Pieces Over 3 ft / DiaInFt	8.447187E-03
AS (Alloy)	Total	No. of Pipe Pieces Over 3 ft / Ton	5.336939
AS (Alloy)	Total	No. of Pipe Pieces over 3 ft / Unit	4.944609E-02
AS (Alloy)	Total	No. of Flanges / Footage	5.53761E-03
AS (Alloy)	Total	No. of Flanges / DiaInFt	9.231899E-04
AS (Alloy)	Total	No. of Flanges / Ton	0.583272
AS (Alloy)	Total	No. of Flanges / Unit	5.403945E-03
AS (Alloy)	Total	No. of In-Line Fittings / Footage	1.107522E-03
AS (Alloy)	Total	No. of In-Line Fittings / DiaInFt	1.84638E-04
AS (Alloy)	Total	No. of In-Line Fittings / Ton	0.1166544
AS (Alloy)	Total	No. of In-Line Fittings / Unit	1.080789E-03
AS (Alloy)	Total	No. of Non-In-Line Fittings / Footage	5.205353E-02
AS (Alloy)	Total	No. of Non-In-Line Fittings / DiaInFt	8.677985E-03
AS (Alloy)	Total	No. of Non-In-Line Fittings / Ton	5.482757
AS (Alloy)	Total	No. of Non-In-Line Fittings / Unit	5.079708E-02
AS (Alloy)	Total	No. of Valves / Footage	0
AS (Alloy)	Total	No. of Valves / DiaInFt	0
AS (Alloy)	Total	No. of Valves / Ton	0
AS (Alloy)	Total	No.of Valves / Unit	0
AS (Alloy)	Total	No. of Supports / Footage	0
AS (Alloy)	Total	No. of Supports / DiaInFt	0
AS (Alloy)	Total	No. of Supports / Ton	0
AS (Alloy)	Total	No. of Supports / Unit	0
AS (Alloy)	Total	No. of Misc. / Footage	5.53761E-04
AS (Alloy)	Total	No. of Misc. / DiaInFt	9.231899E-05
AS (Alloy)	Total	No. of Misc. / Ton	0.0583272
AS (Alloy)	Total	No.of Misc. / Unit	5.403945E-04
AS (Alloy)	Total	Pipe Weight / Spool Weight	0.9404383
AS (Alloy)	Total	Non-Pipe Weight / Spool Weight	5.956174E-02
AS (Alloy)	Total	No. of Connections (Design Welds) / Footage	5.786802E-02

Material	Size	Ratio Description	Ratio
AS (Alloy)	Total	No. of Connections (Design Welds) / DiaInFt	9.647334E-03
AS (Alloy)	Total	No. of Connections (Design Welds) / Ton	6.095192
AS (Alloy)	Total	No. of Connections (Design Welds) / Unit	5.647122E-02
AS (Alloy)	Total	No. of Multi-Station Roll Welds / Footage	2.353484E-02
AS (Alloy)	Total	No. of Multi-Station Roll Welds / DiaInFt	3.923557E-03
AS (Alloy)	Total	No. of Multi-Station Roll Welds / Ton	2.478906
AS (Alloy)	Total	No. of Multi-Station Roll Welds / Unit	2.296676E-02
AS (Alloy)	Total	No. of Repaired Welds / Footage	3.322566E-03
AS (Alloy)	Total	No. of Repaired Welds / DiaInFt	5.539139E-04
AS (Alloy)	Total	No. of Repaired Welds / Ton	0.3499632
AS (Alloy)	Total	No. of Repair Welds / Unit	3.242367E-03
AS (Alloy)	Total	BW DiaIn / Design Weld DiaIn	0.9782972
AS (Alloy)	Total	BW Equiv.DiaIn / Design Weld Equiv.DiaIn	0.9782972
AS (Alloy)	Total	BW Vol. / Design Weld Vol.	0.9940413
AS (Alloy)	Total	SW DiaIn / Design Weld DiaIn	1.335559E-02
AS (Alloy)	Total	SW Equiv.DiaIn / Design Weld Equiv.DiaIn	1.335559E-02
AS (Alloy)	Total	SW Vol. / Design Weld Vol.	3.170402E-03
AS (Alloy)	Total	OL DiaIn / Design Weld DiaIn	8.347246E-03
AS (Alloy)	Total	OL Equiv.DiaIn / Design Weld Equiv.DiaIn	8.347246E-03
AS (Alloy)	Total	OL Vol. / Design Weld Vol.	2.788416E-03
AS (Alloy)	Total	Pressure Attachment / Design Weld DiaIn	0.9866444
AS (Alloy)	Total	Pressure Attachment / Design Weld Equiv.DiaIn	0.9866444
AS (Alloy)	Total	Pressure Attachment / Design Weld Vol.	0.9968296
AS (Alloy)	Total	Non Pressure Attachment / Design Weld DiaIn	1.335559E-02
AS (Alloy)	Total	Non Pressure Attachment / Design Weld Equiv.DiaIn	1.335559E-02
AS (Alloy)	Total	Non Pressure Attachment / Design Weld Vol.	3.170402E-03
AS (Alloy)	Total	Position Weld / Design Weld DiaIn	0
AS (Alloy)	Total	Position Weld / Design Weld Equiv.DiaIn	0
AS (Alloy)	Total	Position Weld / Design Weld Vol.	0
AS (Alloy)	Total	Roll Weld / Design Weld DiaIn	1
AS (Alloy)	Total	Roll Weld / Design Weld Equiv.DiaIn	1
AS (Alloy)	Total	Roll Weld / Design Weld Vol.	1
AS (Alloy)	Total	Multi-Station Roll Weld / Design Weld DiaIn	0.4257095
AS (Alloy)	Total	Multi-Station Roll Weld / Design Weld Equiv.DiaIn	0.4257095
AS (Alloy)	Total	Multi-Station Roll Weld / Design Weld Vol.	0.432653
AS (Alloy)	Total	Single-Station Roll Weld / Design Weld DiaIn	0.5742905
AS (Alloy)	Total	Single-Station Roll Weld / Design Weld Equiv.DiaIn	0.5742905
AS (Alloy)	Total	Single-Station Roll Weld / Design Weld Vol.	0.567347
AS (Alloy)	Total	Tig Process Weld / Design Weld Vol.	0.1067211

Material	Size	Ratio Description	Ratio
AS (Alloy)	Total	Mig Process Weld / Design Weld Vol.	0
AS (Alloy)	Total	FCAW Process Weld / Design Weld Vol.	0
AS (Alloy)	Total	Stick Process Weld / Design Weld Vol.	0.893279
AS (Alloy)	Total	SubArc Process Weld / Design Weld Vol.	0
AS (Alloy)	Total	Rotoweld Process Weld / Design Weld Vol.	0
AS (Alloy)	Total	Repair Rate (No. of R / R+A)	2.933985E-02
AS (Alloy)	Total	No. of Cut Sheet Revision / No. of Spool	0
AS (Alloy)	Total	RT rate /Spool	100
AS (Alloy)	Total	MT rate /Spool	100
AS (Alloy)	Total	PT rate /Spool	0
AS (Alloy)	Total	PMI rate /Spool	100
AS (Alloy)	Total	PWHT rate /Spool	1
AS (Alloy)	Total	FT rate /Spool	0
AS (Alloy)	Total	BHN rate /Spool	100
AS (Alloy)	Total	VT rate /Spool	100
AS (Alloy)	Total	UT rate /Spool	100
AS (Alloy)	Total	Non-Welded Spool/Welded Spool (Weight)	0
AS (Alloy)	Total	Non-Welded Spool/Welded Spool (Units)	0
AS (Alloy)	6-14"	No. of Pipe Pieces / Footage	5.316105E-02
AS (Alloy)	6-14"	No. of Pipe Pieces / DiaInFt	8.862623E-03
AS (Alloy)	6-14"	No. of Pipe Pieces / Ton	5.599411
AS (Alloy)	6-14"	No. of Pipe Pieces / Unit	5.187787E-02
AS (Alloy)	6-14"	No. of Pipe Pieces Over 3 ft / Footage	5.066913E-02
AS (Alloy)	6-14"	No. of Pipe Pieces Over 3 ft / DiaInFt	8.447187E-03
AS (Alloy)	6-14"	No. of Pipe Pieces Over 3 ft / Ton	5.336939
AS (Alloy)	6-14"	No. of Pipe Pieces over 3 ft / Unit	4.944609E-02
AS (Alloy)	6-14"	No. of Flanges / Footage	5.53761E-03
AS (Alloy)	6-14"	No. of Flanges / DiaInFt	9.231899E-04
AS (Alloy)	6-14"	No. of Flanges / Ton	0.583272
AS (Alloy)	6-14"	No. of Flanges / Unit	5.403945E-03
AS (Alloy)	6-14"	No. of In-Line Fittings / Footage	1.107522E-03
AS (Alloy)	6-14"	No. of In-Line Fittings / DiaInFt	1.84638E-04
AS (Alloy)	6-14"	No. of In-Line Fittings / Ton	0.1166544
AS (Alloy)	6-14"	No. of In-Line Filtings / Unit	1.080789E-03
AS (Alloy)	6-14"	No. of Non-In-Line Fittings / Footage	5.205353E-02
AS (Alloy)	6-14"	No. of Non-In-Line Fittings / DiaInFt	8.677985E-03
AS (Alloy)	6-14"	No. of Non-In-Line Fittings / Ton	5.482757
AS (Alloy)	6-14"	No. of Non-In-Line Filtings / Unit	5.079708E-02
AS (Alloy)	6-14"	No. of Valves / Footage	0

Material	Size	Ratio Description	Ratio
AS (Alloy)	6-14"	No. of Valves / DiaInFt	0
AS (Alloy)	6-14"	No. of Valves / Ton	0
AS (Alloy)	6-14"	No.of Valves / Unit	0
AS (Alloy)	6-14"	No. of Supports / Footage	0
AS (Alloy)	6-14"	No. of Supports / DiaInFt	0
AS (Alloy)	6-14"	No. of Supports / Ton	0
AS (Alloy)	6-14"	No. of Supports / Unit	0
AS (Alloy)	6-14"	No. of Misc. / Footage	5.53761E-04
AS (Alloy)	6-14"	No. of Misc. / DiaInFt	9.231899E-05
AS (Alloy)	6-14"	No. of Misc. / Ton	0.0583272
AS (Alloy)	6-14"	No.of Misc. / Unit	5.403945E-04
AS (Alloy)	6-14"	Pipe Weight / Spool Weight	0.9404383
AS (Alloy)	6-14"	Non-Pipe Weight / Spool Weight	5.956174E-02
AS (Alloy)	6-14"	No. of Connections (Design Welds) / Footage	5.786802E-02
AS (Alloy)	6-14"	No. of Connections (Design Welds) / DiaInFt	9.647334E-03
AS (Alloy)	6-14"	No. of Connections (Design Welds) / Ton	6.095192
AS (Alloy)	6-14"	No. of Connections (Design Welds) / Unit	5.647122E-02
AS (Alloy)	6-14"	No. of Multi-Station Roll Welds / Footage	2.353484E-02
AS (Alloy)	6-14"	No. of Multi-Station Roll Welds / DiaInFt	3.923557E-03
AS (Alloy)	6-14"	No. of Multi-Station Roll Welds / Ton	2.478906
AS (Alloy)	6-14"	No. of Multi-Station Roll Welds / Unit	2.296676E-02
AS (Alloy)	6-14"	No. of Repaired Welds / Footage	3.322566E-03
AS (Alloy)	6-14"	No. of Repaired Welds / DiaInFt	5.539139E-04
AS (Alloy)	6-14"	No. of Repaired Welds / Ton	0.3499632
AS (Alloy)	6-14"	No. of Repair Welds / Unit	3.242367E-03
AS (Alloy)	6-14"	BW DiaIn / Design Weld DiaIn	0.9782972
AS (Alloy)	6-14"	BW Equiv.DiaIn / Design Weld Equiv.DiaIn	0.9782972
AS (Alloy)	6-14"	BW Vol. / Design Weld Vol.	0.9940413
AS (Alloy)	6-14"	SW DiaIn / Design Weld DiaIn	1.335559E-02
AS (Alloy)	6-14"	SW Equiv.DiaIn / Design Weld Equiv.DiaIn	1.335559E-02
AS (Alloy)	6-14"	SW Vol. / Design Weld Vol.	3.170402E-03
AS (Alloy)	6-14"	OL DiaIn / Design Weld DiaIn	8.347246E-03
AS (Alloy)	6-14"	OL Equiv.DiaIn / Design Weld Equiv.DiaIn	8.347246E-03
AS (Alloy)	6-14"	OL Vol. / Design Weld Vol.	2.788416E-03
AS (Alloy)	6-14"	Pressure Attachment / Design Weld DiaIn	0.9866444
AS (Alloy)	6-14"	Pressure Attachment / Design Weld Equiv.DiaIn	0.9866444
AS (Alloy)	6-14"	Pressure Attachment / Design Weld Vol.	0.9968296
AS (Alloy)	6-14"	Non Pressure Attachment / Design Weld DiaIn	1.335559E-02
AS (Alloy)	6-14"	Non Pressure Attachment / Design Weld Equiv.DiaIn	1.335559E-02

Material	Size	Ratio Description	Ratio
AS (Alloy)	6-14"	Non Pressure Attachment / Design Weld Vol.	3.170402E-03
AS (Alloy)	6-14"	Position Weld / Design Weld DiaIn	0
AS (Alloy)	6-14"	Position Weld / Design Weld Equiv.DiaIn	0
AS (Alloy)	6-14"	Position Weld / Design Weld Vol.	0
AS (Alloy)	6-14"	Roll Weld / Design Weld DiaIn	1
AS (Alloy)	6-14"	Roll Weld / Design Weld Equiv.DiaIn	1
AS (Alloy)	6-14"	Roll Weld / Design Weld Vol.	1
AS (Alloy)	6-14"	Multi-Station Roll Weld / Design Weld DiaIn	0.4257095
AS (Alloy)	6-14"	Multi-Station Roll Weld / Design Weld Equiv.DiaIn	0.4257095
AS (Alloy)	6-14"	Multi-Station Roll Weld / Design Weld Vol.	0.432653
AS (Alloy)	6-14"	Single-Station Roll Weld / Design Weld DiaIn	0.5742905
AS (Alloy)	6-14"	Single-Station Roll Weld / Design Weld Equiv.DiaIn	0.5742905
AS (Alloy)	6-14"	Single-Station Roll Weld / Design Weld Vol.	0.567347
AS (Alloy)	6-14"	Tig Process Weld / Design Weld Vol.	0.1067211
AS (Alloy)	6-14"	Mig Process Weld / Design Weld Vol.	0
AS (Alloy)	6-14"	FCAW Process Weld / Design Weld Vol.	0
AS (Alloy)	6-14"	Stick Process Weld / Design Weld Vol.	0.893279
AS (Alloy)	6-14"	SubArc Process Weld / Design Weld Vol.	0
AS (Alloy)	6-14"	Rotoweld Process Weld / Design Weld Vol.	0
AS (Alloy)	6-14"	Repair Rate (No. of R / R+A)	2.933985E-02
AS (Alloy)	6-14"	No. of Cut Sheet Revision / No. of Spool	0
AS (Alloy)	6-14"	RT rate /Spool	100
AS (Alloy)	6-14"	MT rate /Spool	100
AS (Alloy)	6-14"	PT rate /Spool	0
AS (Alloy)	6-14"	PMI rate /Spool	100
AS (Alloy)	6-14"	PWHT rate /Spool	1
AS (Alloy)	6-14"	FT rate /Spool	0
AS (Alloy)	6-14"	BHN rate /Spool	100
AS (Alloy)	6-14"	VT rate /Spool	100
AS (Alloy)	6-14"	UT rate /Spool	100
AS (Alloy)	6-14"	Non-Welded Spool/Welded Spool (Weight)	0
AS (Alloy)	6-14"	Non-Welded Spool/Welded Spool (Units)	0

APPENDIX C: USERS' MANUAL FOR FAB_OLAP

FabMaster processed project data individually and warehoused the item-coded project information in an easy-to-access format. Fab_OLAP provides the functionality of viewing and analyzing productivity-related information across all projects that have been processed by FabMaster. Fab_OLAP is an On-line Analytical Processing System custom-developed for the Fabrication Facilities of PCL Industrial Constructors, Inc. The system features dynamic query, graphic presentation, and the functionality of statistical analysis on 105 ratios of labor productivity/spool configuraton complexity/quality control. It is an advanced decision-support tool for management to grasp the trend in the historical project data and identify exceptional problems in the work at hand.

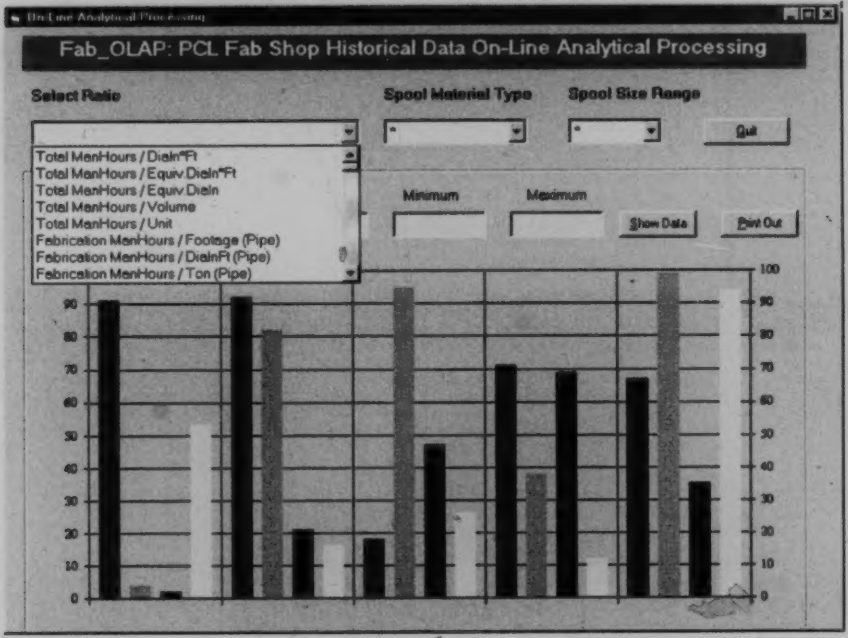

On-Line Analytical Processing
Fab_OLAP: PCL Fab Shop Historical Data On-Line Analytical Processing
Select Ratio
Spool Material Type
Spool Size Range
Quit
Total ManHours / DiaIn*Ft
Total ManHours / Equiv.DiaIn*Ft
Total ManHours / Equiv.DiaIn
Total ManHours / Volume
Total ManHours / Unit
Fabrication ManHours / Footage (Pipe)
Fabrication ManHours / DiaInFt (Pipe)
Fabrication ManHours / Ton (Pipe)
Minimum
Maximum
Show Data
Print Out
0
10
20
30
40
50
60
70
80
90
100

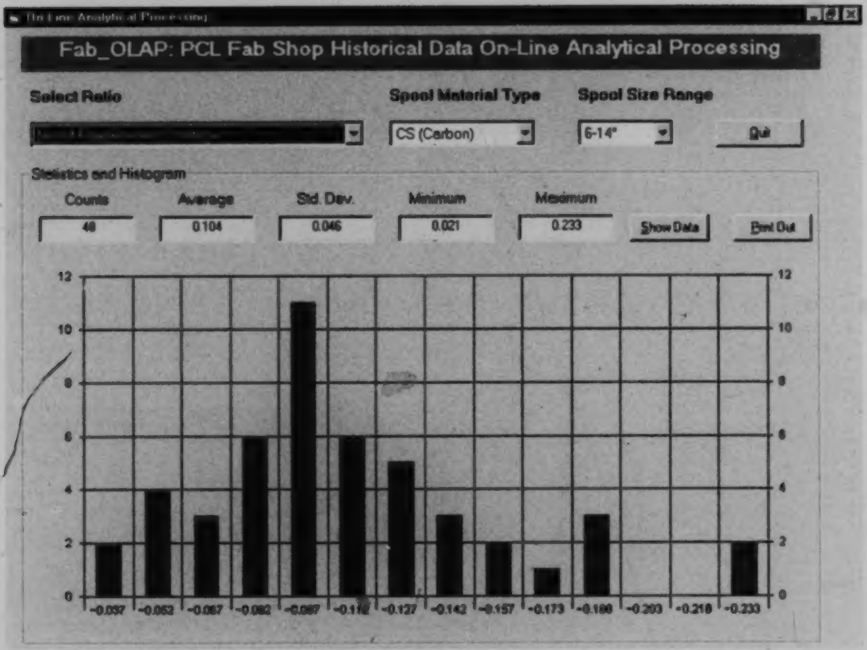
Figure C-1: Select one ratio

Step 1. Load the program and select one ratio

User selects one ratio from the "Select Ratio" dropdown list, which includes all the 105 ratios computed in FabMaster, as shown in Figure C-1.

Step 2. Apply Filters on Material Type and Size Range

Fab_OLAP uses the standard codes of the company for the material types and size ranges. Fab_OLAP helps user explore data in decision-oriented ways and allows user to view data and get at them form different perspectives along the dimension of



On Line Analytical Processing
Fab_OLAP: PCL Fab Shop Historical Data On-Line Analytical Processing
Quit
Statistics and Histogram
Show Data
Print Out
12
10
8
6
4
2
0
-0.037
-0.052
-0.067
-0.082
-0.097
-0.112
-0.127
-0.142
-0.157
-0.173
-0.188
-0.203
-0.218
-0.233

Spool Material Type: CS (Carbon)

Spool Size Range: 6-14"

Counts: 48

Average: 0.104

Std. Dev.: 0.046

Minimum: 0.021

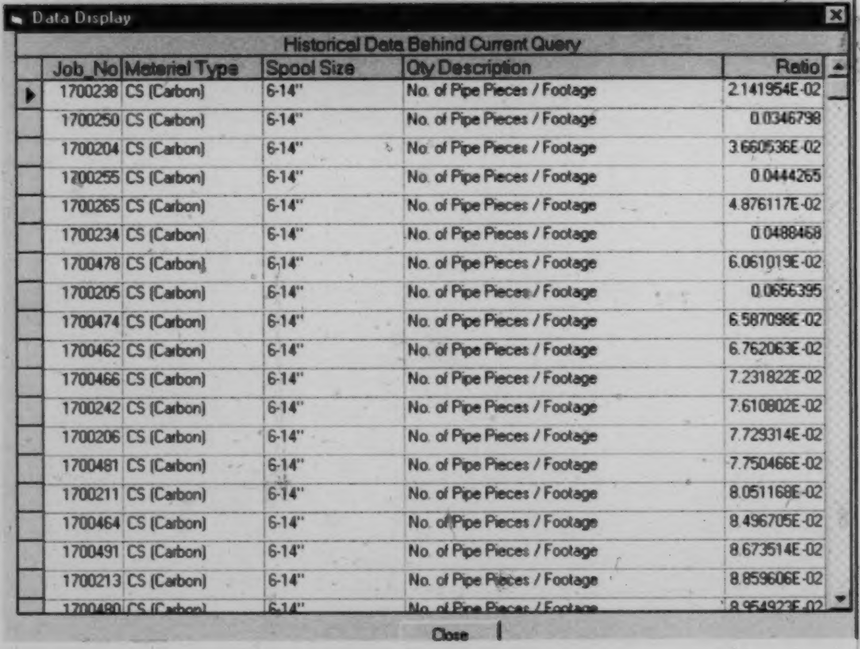
Maximum: 0.233

Figure C-2: Trial on "number of pipe pieces per foot"

material and size. The histogram along with statistical analysis results for the selected ratio is presented on screen and updated automatically. Figure C-2 shows the trial based on "carbon steel, 6-14 inch spool, number of pipe pieces per foot of pipe".

Step 3. Drill into details of data

Fab_OLAP allows user to drill down to details of data by clicking the "View Data" button. Figure C-3 shows the data behind the selected ratio.



Historical Data Behind Current Query

Job_No	Material Type	Spool Size	Qty Description	Ratio
1700238	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	2.141954E-02
1700250	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	0.0346798
1700204	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	3.660536E-02
1700255	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	0.0444265
1700265	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	4.876117E-02
1700234	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	0.0488468
1700478	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	6.061019E-02
1700205	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	0.0656395
1700474	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	6.587098E-02
1700462	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	6.762063E-02
1700466	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	7.231822E-02
1700242	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	7.610802E-02
1700206	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	7.729314E-02
1700481	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	7.750466E-02
1700211	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	8.051168E-02
1700464	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	8.496705E-02
1700491	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	8.673514E-02
1700213	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	8.859606E-02
1700480	CS (Carbon)	6-14"	No. of Pipe Pieces / Footage	8.954923E-02

Figure C-3: View details of data

Step 4. Print out the trial and statistical analysis results

User clicks the "Print out" button to print a hard copy of the current trial and statistical analysis results including the histogram for record.

APPENDIX D: USER'S MANUAL FOR PIPINGMASTER

PipingMaster is a historical project data warehousing system customized for the field construction systems of PCL Industrial Constructors, Inc. It is an automated data processing tool to extract raw data from Labor Cost Control System, Estimating System, and Quality Control System, and convert raw data into useful productivity information based on embedded expert rules. Pipe handling and welding are processed by PipingMaster independently, but in similar fashions including the user interfaces and program logic. Thus, Pipe handling is selected to illustrate the program flow in the following steps.

Step 1. Import raw data in standard format

User imports three raw data tables for each project into the database manually to allow for PipingMaster to calculate the quantity of piping work, namely, RD_Project#Hand table for pipe handling, RD_Project#Detl for pipe work components, RD_Project#Weld for pipe welding. The table structures are shown in Figure D-1.

RD_Proj#Hand

Project #
Nominal Size
Schedule
Classification
Material Type
Length (ft)
EstUnitMH

RD_Proj#Detl

Project #
Detl Type
Nominal Size
Classification
Material Type
Quantity
EstUnitMH

RD_Proj#Weld

Project #
Nominal Size
Schedule
Joint Type
Classification
Material Type
# Welds
EstUnitMH

Figure D-1: Structures of Raw Data Tables for A Project

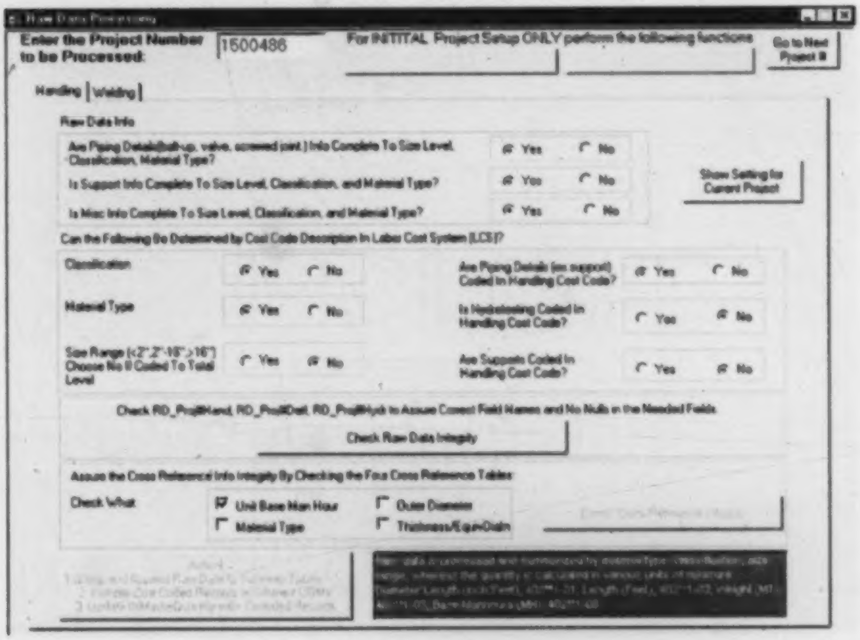
The detailed quantity take-off (in footage) for pipe handling of one project is available in the project estimate only. Usually information is known and complete on the size, the thickness, the material type, and the location classification of each individual pipe section.

The detailed quantity take-off in number of welds for pipe welding of one project is available either in the project estimates or in the field quality control system. In most cases the pipe size, pipe thickness, pipe material type, location classification and weld joint type are known for each individual weld.

Installation of other piping work components (or piping details) includes pipe supports, bolt-ups, valves, screw joints, and miscellaneous items like flanges, specialties, elbows, cuts and bevels. The number and type of work components and estimated unit man-hours for one project are available in the project estimates. However, information on the size, material type, location classification may not be found in the estimate. Therefore, we need to check the raw data integrity of the piping work components prior to processing.

Step 2 Raw Data Integrity Check

The raw data integrity check is controlled by the entered project setting regarding the raw data integrity and methods of actual man-hour cost coding as shown in Figure D-2. User



Enter the Project Number to be Processed: 1500486

For INITITAL Project Setup ONLY perform the following functions

Raw Data Info

Are Piping Details(bolt-up, valve, screwed joint.) Info Complete To Size Level, Classification, Material Type? ● Yes ○ No

Is Support Info Complete To Size Level, Classification, and Material Type? ● Yes ○ No

Is Misc Info Complete To Size Level, Classification, and Material Type? ● Yes ○ No

Can the Following Be Determined by Cost Code Description In Labor Cost System (LCS)?

Classification ● Yes ○ No

Material Type ● Yes ○ No

Size Range (<2",2"-16",>16") Choose No If Coded To Total Level ○ Yes ● No

Are Piping Details (ex support) Coded In Handling Cost Code? ● Yes ○ No

Is Hydrotesting Coded In Handling Cost Code? ○ Yes ● No

Are Supports Coded In Handling Cost Code? ○ Yes ● No

Check RD_ProjIHand, RD_ProjIDet, RD_ProjIHydr to Assure Correct Field Names and No Nulls in the Needed Fields

Assure the Cross Reference Info Integrity By Checking the Four Cross Reference Tables

Check What: ☑ Unit Base Man Hour ☐ Outer Diameter ☐ Material Type ☐ Thickness/EquivDiam

Figure D-2: Main user interface of FabMaster

enters the project number to be processed and answer a number of Yes/No questions about the project. Next, user clicks "Check Raw Data Integrity" button to start the program. User will be prompted to correct any problems due to failure to pass the checks.

The PipingMaster is capable of identifying missing data or incorrect data in the raw data tables. For example, if actual labor hours in the labor cost system were tracked to the level of various classifications of location, then a null in the "Classification" field of the raw data tables will be detected as invalid data and must be corrected for further processing. Three valid types of weld joint, i.e. BW (Butt Weld), SW (Socket Weld), OL (Olet Weld) and five valid types of piping work components are allowed in the RD_Project#Detl table, i.e. bolt-up, valve, screw joint, support, and misc.

Step 3 Check cross-reference Data for Quantity Calculation

User first chooses one of four options and then click the "Check Cross Reference Integrity" button to perform the check for the selected option. Four options should be checked through one by one. User will be prompted to correct raw data or update cross reference tables in case PipingMaster finds a problem. The "Action" button will only be activated when all the raw data checks and cross reference checks are passed.

In PipingMaster, a number of cross-reference tables are involved to calculate the quantity in various units of measurement, i.e. five units of measurement for pipe handling: Diameter*Length (Inch*Feet), Equivalent Diameter*Length (Inch*Feet), Length (Feet), Weight (MT), and Base Manhours (MH); five units of measurement for pipe welding: Diameter (Inch), Equivalent Diameter (Inch), Volume (Cubic Inch), Volume/Thickness (Square Inch), and Base Manhours (MH). Cross-reference integrity check is performed to ascertain that each record in the raw data table can find the needed information in the corresponding cross-reference tables so as to calculate accurate quantities. The formulae used are commonly found in an industrial manual or piping handbook. The relationships between raw data tables and cross-reference tables are shown in Figure D-3 and Figure D-4.

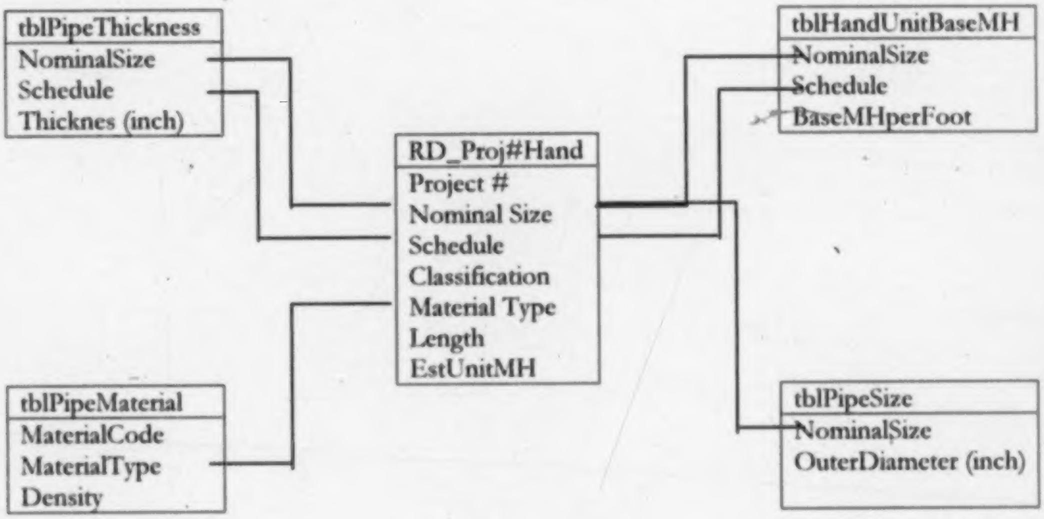

tblPipeThickness
NominalSize
Schedule
Thicknes (inch)
tblHandUnitBaseMH
NominalSize
Schedule
BaseMHperFoot
RD_Proj#Hand
Project #
Nominal Size
Schedule
Classification
Material Type
Length
EstUnitMH
tblPipeMaterial
MaterialCode
MaterialType
Density
tblPipeSize
NominalSize
OuterDiameter (inch)

Figure D-3: Handling: X-Reference Information Integrity Check

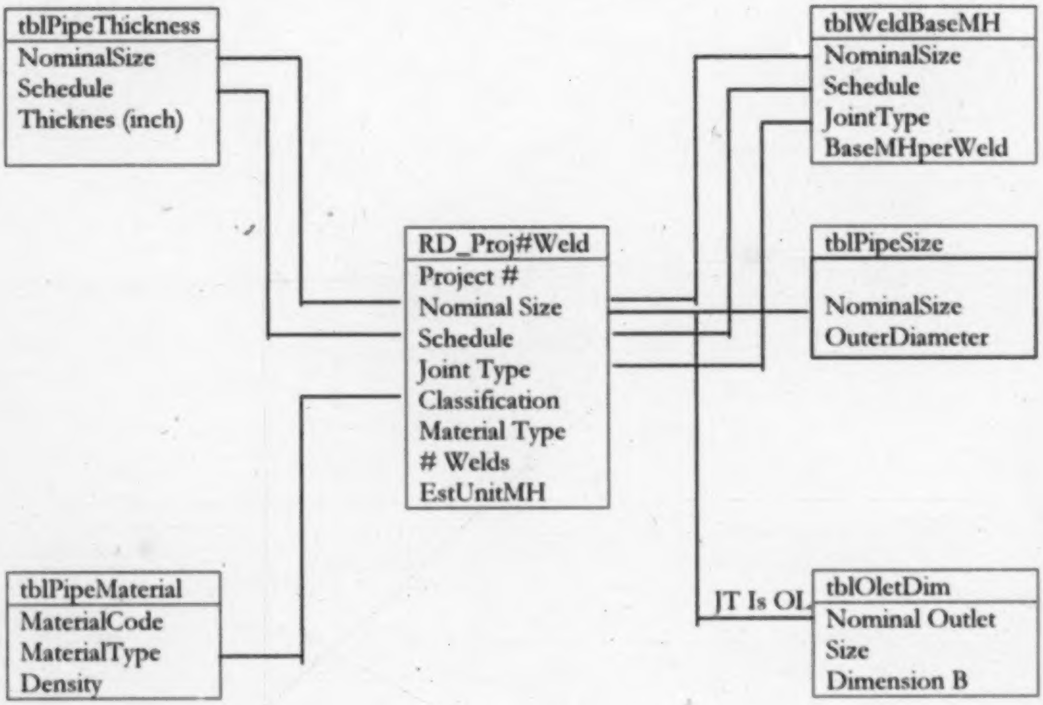

tblPipeThickness
NominalSize
Schedule
Thicknes (inch)
tblWeldBaseMH
NominalSize
Schedule
JointType
BaseMHperWeld
RD_Proj#Weld
Project #
Nominal Size
Schedule
Joint Type
Classification
Material Type
# Welds
EstUnitMH
tblPipeSize
NominalSize
OuterDiameter
tblPipeMaterial
MaterialCode
MaterialType
Density
JT Is OL
tblOletDim
Nominal Outlet Size
Dimension B

Figure D-4: Welding: X-Reference Information Integrity Check

Step 4 Generate Aggregate Cost Codes and Calculate Quantities

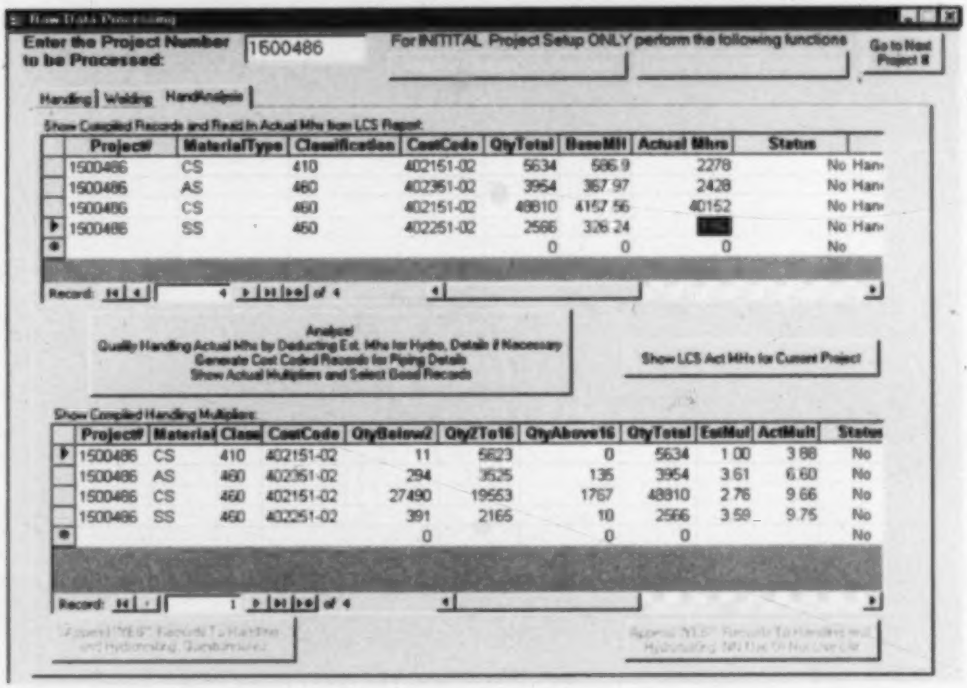
User hits the "Action" button to generate aggregate cost codes to the level of project number, classification of location, material type, size range, activity, and unit of measure. The total quantities and quantities breakdown for size ranges, along with the generated cost codes will be appended a summary table called "tblQuantityMaster". Table 1 shows sample records in the summary table for one relatively small job.

Table D-1: Sample of Quantity Calculation Summary Table in PipingMaster

Project#	Material	Class	CostCode	Qty<2	Qty2-16	Qty>16	QtyTotal	Description
1500486	CS	410	402152-02	2.17	2717.33	0	2719.50	Welding Total Volume/Thickness
1500486	CS	410	402151-02	11	5623	0	5634	Handling Total Feet
1500486	CS	460	402151-02	27490	19553	1767	48810	Handling Total Feet
1500486	CS	460	402152-02	11839.85	7076.22	1763.26	20679.32	Welding Total Volume/Thickness
1500486	AS	460	402351-02	294	3525	135	3954	Handling Total Feet
1500486	AS	460	402352-02	6.59	495.49	85.71	587.80	Welding Total Volume/Thickness
1500486	SS	460	402251-02	391	2165	10	2566	Handling Total Feet
1500486	SS	460	402252-02	11.96	767.77	0	779.74	Welding Total Volume/Thickness
1500486	TOT	410	402051-02	11	5623	0	5634	Hand Tot Matl Tot Size Ft
1500486	TOT	460	402051-02	28175	25243	1912	55330	Hand Tot Matl Tot Size Ft

Step 5 Enter Actual Hours and Compute Actual Degree-of-difficulty Factors

Following generating the cost codes and calculating the quantities, PipingMaster shifts focus to productivity analysis page as shown in Figure D-5. User reads actual manhours and enters into the "actual manhours" column for corresponding records. Next, user hits the "Analyze" button to let PipingMaster figure out the actual labor hours for pipe handling based on the project setting about actual labor cost tracking practice and the embedded expert rules for handling different scenarios. Eventually, the actual degree-of-difficulty factors are computed for each record and listed against the factors estimators have used for comparison. After comparison, user decides on which records are valid for NN to use by switching the status of one record from No to Yes.


For INITITAL Project Setup ONLY perform the following functions
Go to Next Project #
Handling
Welding
HandAnalysis
Analyze!
Qualify Handling Actual Mhs by Deducting Est. Mhs for Hydro, Details if Necessary
Generate Cost Coded Records for Piping Details
Show Actual Multipliers and Select Good Records
Show LCS Act MHs for Current Project

Enter the Project Number to be Processed: 1500486

Show Compiled Records and Read In Actual Mhs from LCS Report:

Project#	MaterialType	Classification	CostCode	QtyTotal	BaseMH	Actual Mhrs	Status
1500486	CS	410	402151-02	5634	586.9	2278	No Han
1500486	AS	460	402351-02	3954	367.97	2428	No Han
1500486	CS	460	402151-02	48810	4157.56	40152	No Han
1500486	SS	460	402251-02	2566	326.24	[illegible]	No Han
				0	0	0	No

Record: of 4

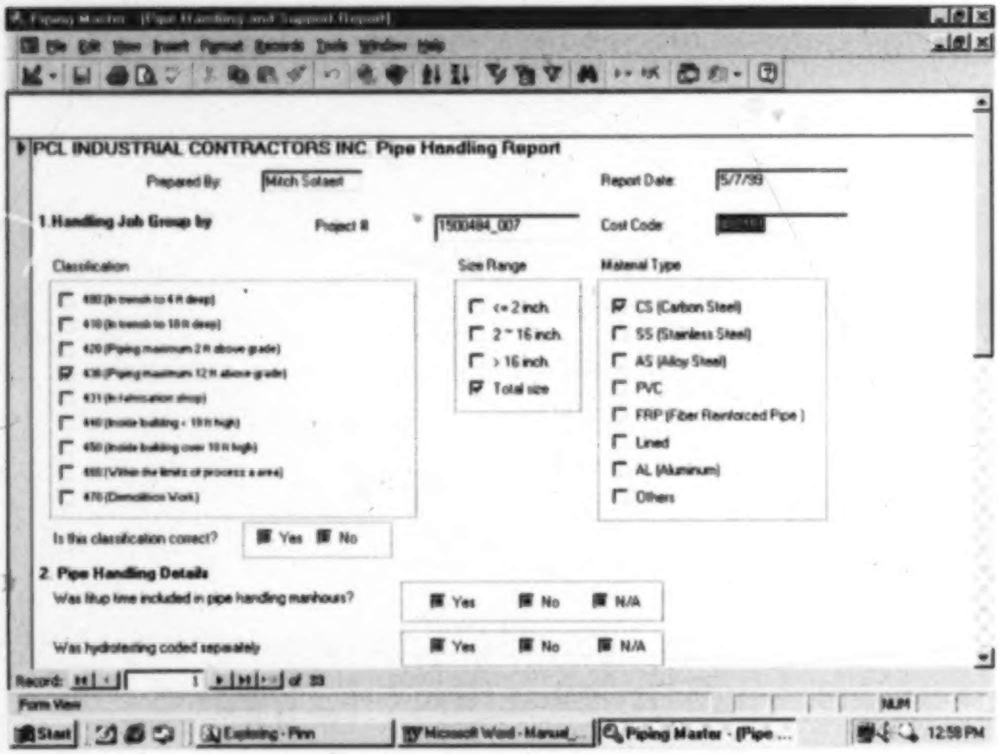
Show Compiled Handling Multipliers:

Project#	Material	Class	CostCode	QtyBelow2	Qty2To16	QtyAbove16	QtyTotal	EstMul	ActMult	Statu
1500486	CS	410	402151-02	11	5623	0	5634	1.00	3.88	No
1500486	AS	460	402351-02	294	3525	135	3954	3.61	6.60	No
1500486	CS	460	402151-02	27490	19553	1767	48810	2.76	9.66	No
1500486	SS	460	402251-02	391	2165	10	2566	3.59	9.75	No
				0		0	0			No

Record: 1 of 4

Figure D-5: Productivity Analysis Page for Pipe Handling

Step 6 Make Questionnaires for Valid Records


Piping Master - [Pipe Handling and Support Report]
File Edit View Insert Format Records Tools Window Help
Form View
NUM
Start
Exploring - Pim
Microsoft Word - Manual_...
Piping Master - [Pipe ...
12:58 PM

PCL INDUSTRIAL CONTRACTORS INC Pipe Handling Report

Prepared By: Mitch Solaert

Report Date: 5/7/99

1. Handling Job Group by

Project #: 1500484_007

Cost Code: [illegible]

Classification
- ☐ 400 (In trench to 4 ft deep)
- ☐ 410 (In trench to 10 ft deep)
- ☐ 420 (Piping maximum 2 ft above grade)
- ☑ 430 (Piping maximum 12 ft above grade)
- ☐ 431 (In fabrication shop)
- ☐ 440 (Inside building < 10 ft high)
- ☐ 450 (Inside building over 10 ft high)
- ☐ 460 (Within the limits of process a area)
- ☐ 470 (Demolition Work)

Size Range
- ☐ <= 2 inch
- ☐ 2 ~ 16 inch
- ☐ > 16 inch
- ☑ Total size

Material Type
- ☑ CS (Carbon Steel)
- ☐ SS (Stainless Steel)
- ☐ AS (Alloy Steel)
- ☐ PVC
- ☐ FRP (Fiber Reinforced Pipe)
- ☐ Lined
- ☐ AL (Aluminum)
- ☐ Others

Is this classification correct? ■ Yes ■ No

2. Pipe Handling Details

Was fitup time included in pipe handing manhours? ■ Yes ■ No ■ N/A

Was hydrotesting coded sepaately ■ Yes ■ No ■ N/A

Record: 1 of 33

Figure D-6: Sample of Pipe Handling Questionnaire

PipingMaster makes questionnaires for those valid records as confirmed by user. Figure D-6 shows a sample questionnaire.

Following the above six steps, PipingMaster processes one project and convert raw data into accurate cost-coded productivity information for further productivity analysis. Figure D-7 shows the flow chart of the whole program.

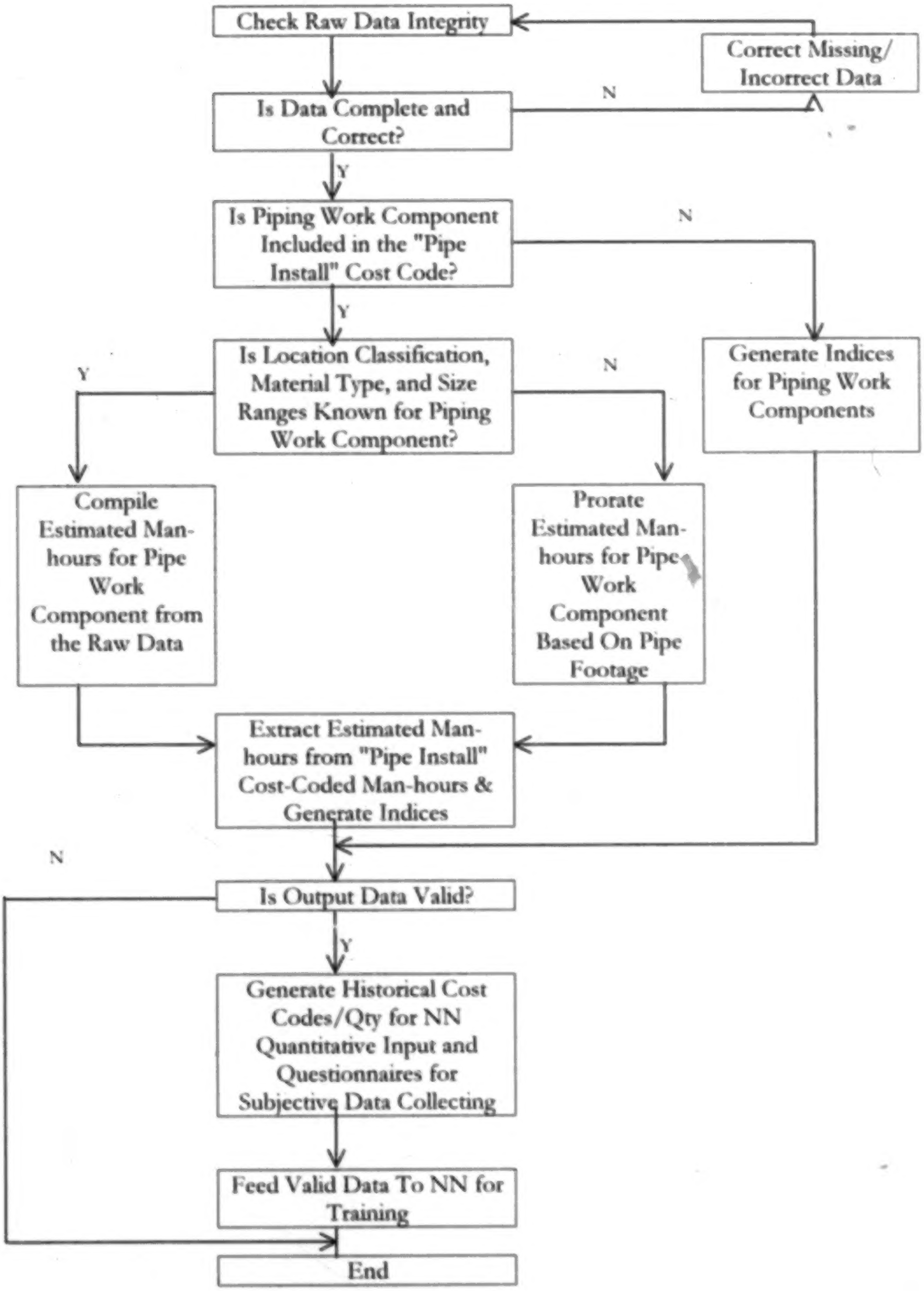

Check Raw Data Integrity
Correct Missing/ Incorrect Data
N
Is Data Complete and Correct?
Y
Is Piping Work Component Included in the "Pipe Install" Cost Code?
N
Y
Is Location Classification, Material Type, and Size Ranges Known for Piping Work Component?
Y
N
Generate Indices for Piping Work Components
Compile Estimated Man-hours for Pipe Work Component from the Raw Data
Prorate Estimated Man-hours for Pipe Work Component Based On Pipe Footage
Extract Estimated Man-hours from "Pipe Install" Cost-Coded Man-hours & Generate Indices
N
Is Output Data Valid?
Y
Generate Historical Cost Codes/Qty for NN Quantitative Input and Questionnaires for Subjective Data Collecting
Feed Valid Data To NN for Training
End

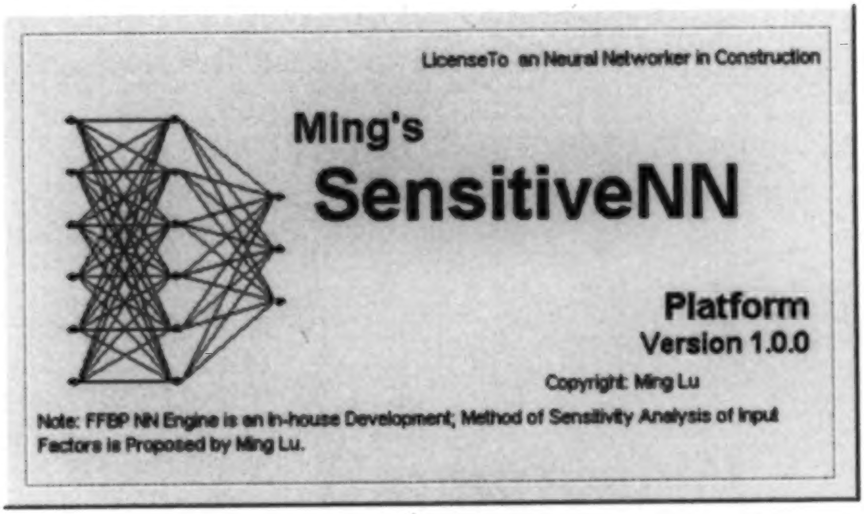
Figure D-7: Program Flow Chart of PipingMaster

APPENDIX E: USER'S MANUAL FOR SENSITIVENN

SensitiveNN is a back propagation Neural-Network based system to analyze the sensitivity of input factors in some complex engineering and management problems that are not amenable to analysis using conventional mathematical models. The sensitivity analysis method is proposed in this thesis.

Step 1 Prepare data for SensitiveNN

The last column in a data table must be named as "Status", which flags the training/testing status for each record. Status 1 stands for a training record, and Status 2 for a testing record, and Status 0 for an ignored record. The next-to-last N columns in a data table contain Actual Output Values of the target risky variables such as actual production rates, N being the number of outputs. All the remaining columns in a data table will be the input factors. There are no requirements imposed on the column names. The trainer will count the number of inputs and outputs according to user's setup of the network, which is discussed in Step 3.



LicenseTo an Neural Networker in Construction

Ming's

SensitiveNN

Platform

Version 1.0.0

Copyright: Ming Lu

Note: FFBP NN Engine is an in-house Development; Method of Sensitivity Analysis of Input Factors is Proposed by Ming Lu.

Figure E-1: Splash Screen of SensitiveNN program

The prepared data table for PINN must be imported to the database file "FFBPNN.mdb" prior to analysis, which is installed with the program and by default under the program folder.

Once the data table is imported, user can start up the program "SensitiveNN.exe". The splash screen shows up as in Figure E-1. By hitting the form, user proceeds to the next step.

Step 2 Link SensitiveNN to data tables and select the one for analysis

Figure E-2 shows the switchboard of the program. User needs to link to "FFBPNN.mdb" first in order to load up data. By clicking the "Link FFBPNN.mdb" button, an "Open File" dialog form pops up as show in Figure E-3.

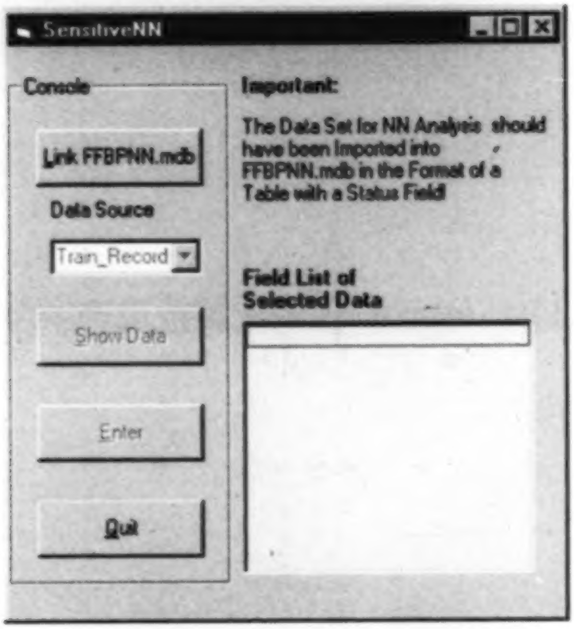
NOTE TO USERS

Page(s) missing in number only; text follows. Page(s) were microfilmed as received.

196

This reproduction is the best copy available.

UMI


SensitiveNN
Console
Link FFBPNN.mdb
Show Data
Enter
Quit

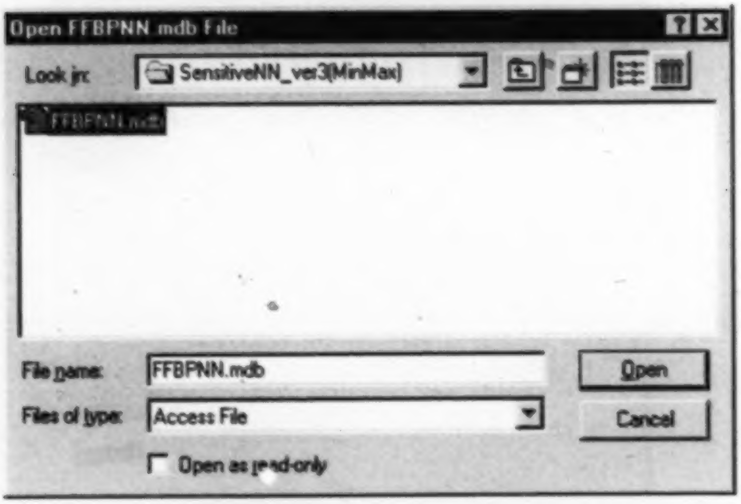
Important:

The Data Set for NN Analysis should have been Imported into FFBPNN.mdb in the Format of a Table with a Status Field!

Data Source: Train_Record

Field List of Selected Data

Figure E-2: Program Switchboard


Open FFBPNN.mdb File
FFBPNN.mdb
Open
Cancel

Look in: SensitiveNN_ver3(MinMax)

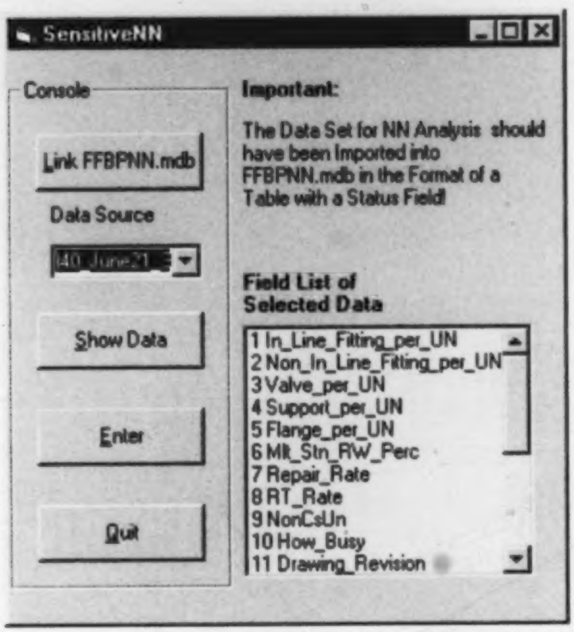
File name: FFBPNN.mdb

Files of type: Access File

☐ Open as read-only

Figure E-3: Open FFBPNN.mdb First

Once FFBPNN.mdb is linked, all the data tables are listed in a combo box for user to select the one for analysis. All the fields in the selected table are numbered and listed in the list box captioned "Field List of Selected Data", as shown in Figure E-4. By clicking the "Show Data" button, user can examine the details of data and edit the train/test status for each record, which is shown in Figure E-5.



Important:

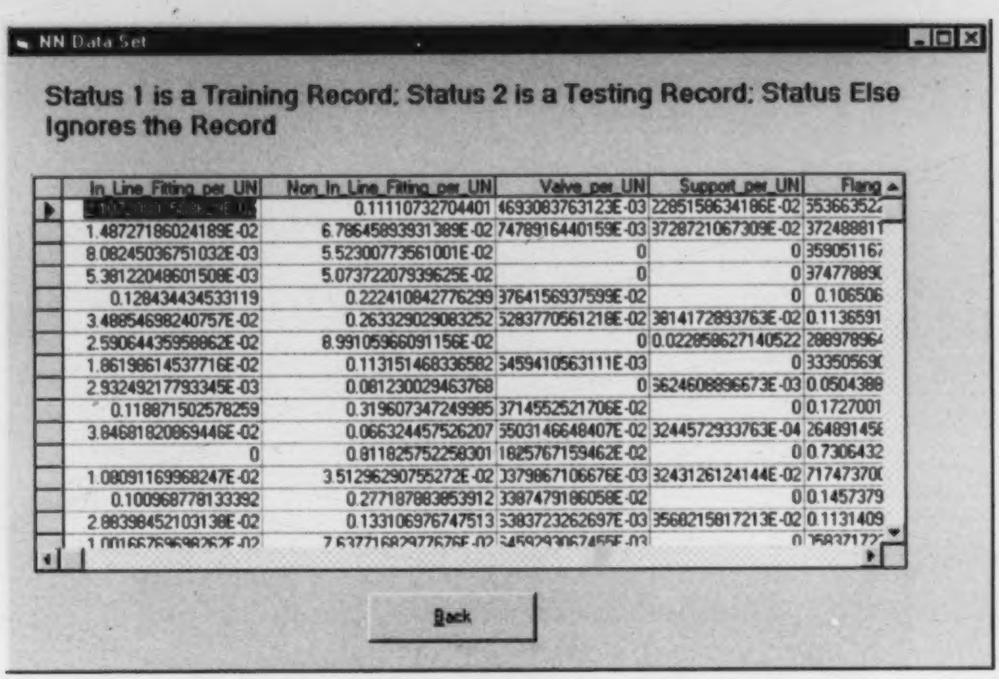
The Date Set for NN Analysis should have been Imported into FFBPNN.mdb in the Format of a Table with a Status Field!

Data Source: 40_June21

Field List of Selected Data

1 In_Line_Fitting_per_UN
2 Non_In_Line_Fitting_per_UN
3 Valve_per_UN
4 Support_per_UN
5 Flange_per_UN
6 Mlt_Stn_RW_Perc
7 Repair_Rate
8 RT_Rate
9 NonCsUn
10 How_Busy
11 Drawing_Revision

Figure E-4: Select data source table



Status 1 is a Training Record; Status 2 is a Testing Record; Status Else Ignores the Record

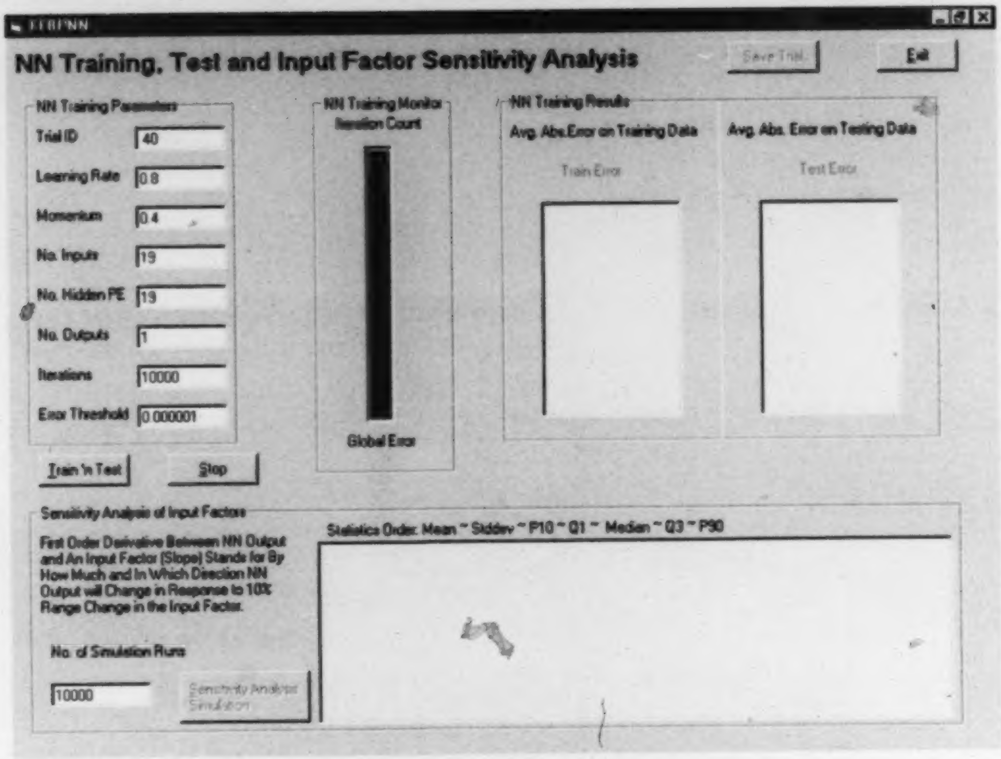
In_Line_Fitting_per_UN	Non_In_Line_Fitting_per_UN	Valve_per_UN	Support_per_UN	Flang
[illegible]	0.111110732704401	4693083763123E-03	2285158634186E-02	353663522
1.48727186024189E-02	6.78645893931389E-02	7478916440159E-03	3728721067309E-02	372488811
8.08245036751032E-03	5.52300773561001E-02	0	0	359051167
5.38122048601508E-03	5.07372207939625E-02	0	0	374778890
0.128434434533119	0.222410842776299	3764156937599E-02	0	0.106506
3.48854698240757E-02	0.263329029083252	5283770561218E-02	3814172893763E-02	0.1136591
2.59064435958862E-02	8.99105966091156E-02	0	0.022858627140522	288978964
1.86198614537716E-02	0.113151468336582	5459410563111E-03	0	333505690
2.93249217793345E-03	0.081230029463768	0	5624608896673E-03	0.0504388
0.118871502578259	0.319607347249985	3714552521706E-02	0	0.1727001
3.84681820869446E-02	0.066324457526207	5503146648407E-02	3244572933763E-04	264891458
0	0.811825752258301	1625767159462E-02	0	0.7306432
1.08091169968247E-02	3.51296290755272E-02	3379867106676E-03	3243126124144E-02	717473700
0.100968778133392	0.277187883853912	3387479186058E-02	0	0.1457379
2.88398452103138E-02	0.133106976747513	5383723262697E-03	3568215817213E-02	0.1131409
1.00166769698262E-02	7.63771682977676E-02	3459293067455E-03	0	158371172

Figure E-5: Examine details of data and edit record status

Step 3 Set up NN structure and learning parameters to train-test BP NN

Following linking to the data source, user click the "Enter" button on the switchboard to enter the main interface of the program, as shown in Figure E-6.

User enters the trial ID, the learning rate, the momentum rate, the number of inputs, the number of hidden processing elements in the middle layer, the number of outputs, the training iterations, and the threshold of global error to terminate learning. It is important to match the number of inputs and outputs with the number of columns of the linked data table in the previous step. User may revert to the switchboard (Figure E-4) anytime for double check by clicking the "Exit" button, and clicks the "Enter" button on the switchboard to restore the main interface. The trial ID is used to identify a specific train-test trial and store the network parameters and weights.


FFBPNN
NN Training, Test and Input Factor Sensitivity Analysis
Save Trial
Exit
NN Training Monitor
Iteration Count
Global Error
NN Training Results
Avg. Abs.Error on Training Data
Avg. Abs. Error on Testing Data
Train Error
Test Error
Train 'n Test
Stop
Sensitivity Analysis Simulation

NN Training Parameters

Trial ID 40
Learning Rate 0.8
Momentum 0.4
No. Inputs 19
No. Hidden PE 19
No. Outputs 1
Iterations 10000
Error Threshold 0.000001

Sensitivity Analysis of Input Factors

First Order Derivative Between NN Output and An Input Factor (Slope) Stands for By How Much and In Which Direction NN Output will Change in Response to 10% Range Change in the Input Factor.

No. of Simulation Runs

10000

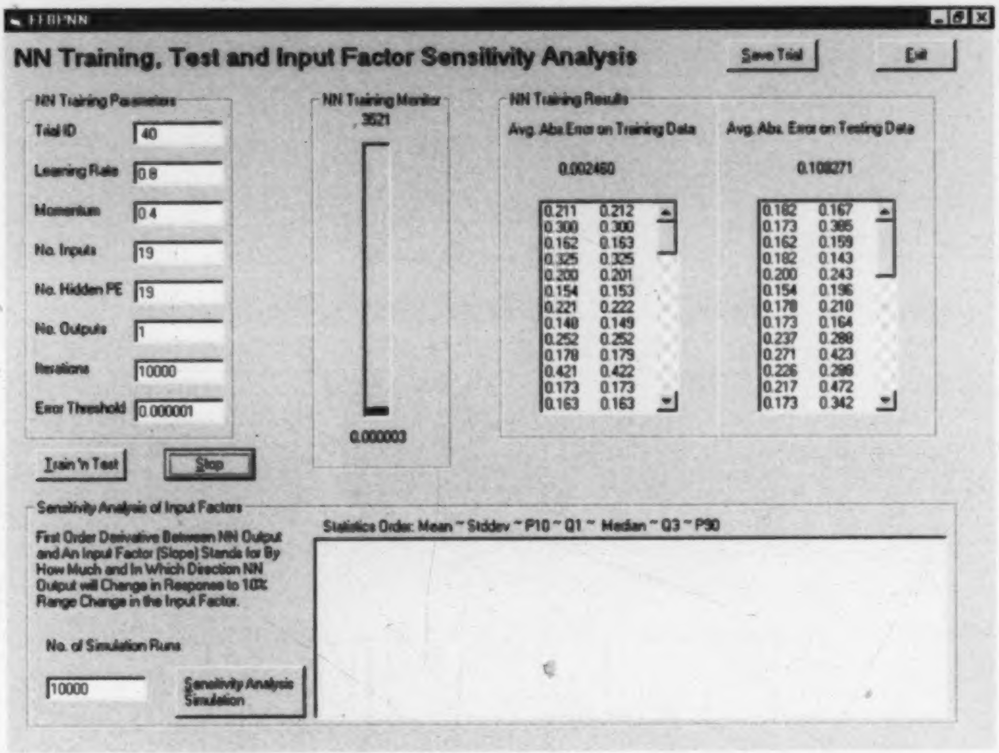
Statistics Order: Mean ~ Stddev ~ P10 ~ Q1 ~ Median ~ Q3 ~ P90

Figure E-6: Program main interface of SensitiveNN

User may refer to the pertinent paper for details of those NN parameters.

Once the network is set up, click the "Train n Test" button to start the learning process, which can be monitored through a progress bar. The current iteration and global error are also shown at the top and bottom of the progress bar dynamically during the learning.

Step 4 Training terminates and investigate learning results


FFBPNN
NN Training, Test and Input Factor Sensitivity Analysis
Save Trial
Exit
Train 'n Test
Stop
Sensitivity Analysis Simulation

NN Training Parameters

Trial ID: 40
Learning Rate: 0.8
Momentum: 0.4
No. Inputs: 19
No. Hidden PE: 19
No. Outputs: 1
Iterations: 10000
Error Threshold: 0.000001

NN Training Monitor

```
3521

0.000003
```

NN Training Results

Avg. Abs Error on Training Data: 0.002460

```
0.211  0.212
0.300  0.300
0.162  0.163
0.325  0.325
0.200  0.201
0.154  0.153
0.221  0.222
0.148  0.149
0.252  0.252
0.178  0.179
0.421  0.422
0.173  0.173
0.163  0.163
```

Avg. Abs. Error on Testing Data: 0.108271

```
0.182  0.167
0.173  0.385
0.162  0.159
0.182  0.143
0.200  0.243
0.154  0.196
0.178  0.210
0.173  0.164
0.237  0.288
0.271  0.423
0.226  0.298
0.217  0.472
0.173  0.342
```

Sensitivity Analysis of Input Factors

First Order Derivative Between NN Output and An Input Factor (Slope) Stands for By How Much and In Which Direction NN Output will Change in Response to 10% Range Change in the Input Factor.

No. of Simulation Runs: 10000

Statistics Order: Mean ~ Stddev ~ P10 ~ Q1 ~ Median ~ Q3 ~ P90

Figure E-7: Check learning results when NN training terminates

The training process terminates when any of the following conditions is satisfied:

1. The current training iteration hits the user-specified total iterations.
2. The current global error is lower than the user-specified threshold of global error.
3. User hits the "Stop" button.

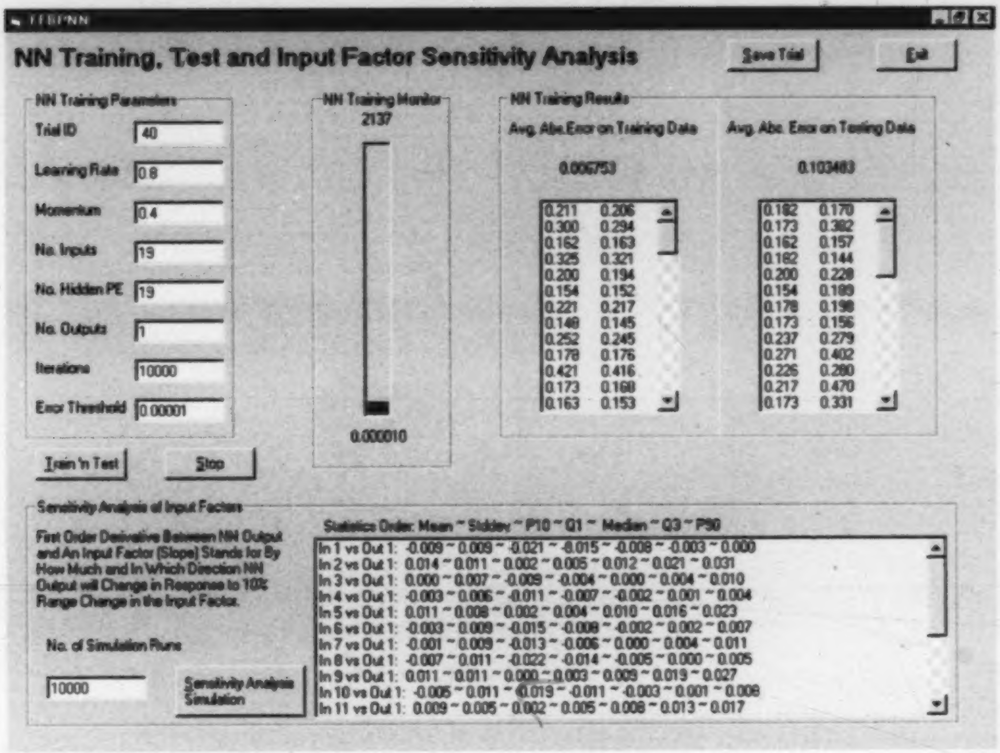
When training terminates, user investigates the learning results by checking final global error and comparing the actual outputs against the NN outputs for both training and testing data as shown in Figure E-7. Note that the average absolute errors for both training data and learning data are computed and shown in the screen as well. If the average absolute errors for both training data and learning data are reasonably small, the network is declared to be trained and the program flow moves to the next step. Otherwise, repeat step 3.

Step 5 Perform Monte Carlo simulation to analyze the sensitivity of input factors

Based on a mature network obtained from Step 4, user specifies the total number of simulation runs in the left lower corner of the main interface and clicks the "Sensitivity Analysis Simulation" button. When the simulation is done, the statistical analysis results of simulation about the input sensitivity between each input-output pair are shown on screen as in Figure E-8. A tab-delimited text file called "SenNNFile.txt" is also generated in the program folder recording the simulation results, which can be imported to Excel for plotting. Note that the text file will be erased next time the simulation is performed, thus user should back it up if needed.

Step 6 Save a trial

User can save a trial including the network structure, learning parameters and final weights of trained network by clicking the "Save Trial" button. The trial ID will be the key for access the network at later times, hence must be remembered.



NN Training, Test and Input Factor Sensitivity Analysis

NN Training Parameters

Trial ID: 40
Learning Rate: 0.8
Momentum: 0.4
No. Inputs: 19
No. Hidden PE: 19
No. Outputs 1
Iterations: 10000
Error Threshold: 0.00001

NN Training Monitor

2137

0.000010

NN Training Results

Avg. Abs.Error on Training Data

0.006753

```
0.211   0.206
0.300   0.294
0.162   0.163
0.325   0.321
0.200   0.194
0.154   0.152
0.221   0.217
0.148   0.145
0.252   0.245
0.178   0.176
0.421   0.416
0.173   0.168
0.163   0.153
```

Avg. Abs. Error on Testing Data

0.103483

```
0.182   0.170
0.173   0.382
0.162   0.157
0.182   0.144
0.200   0.228
0.154   0.189
0.178   0.198
0.173   0.156
0.237   0.279
0.271   0.402
0.226   0.280
0.217   0.470
0.173   0.331
```

Sensitivity Analysis of Input Factors

First Order Derivative Between NN Output and An Input Factor (Slope) Stands for By How Much and In Which Direction NN Output will Change in Response to 10% Range Change in the Input Factor.

No. of Simulation Runs: 10000

Statistics Order: Mean ~ Stddev ~ P10 ~ Q1 ~ Median ~ Q3 ~ P90

```
In 1 vs Out 1:  -0.009 ~ 0.009 ~ -0.021 ~ -0.015 ~ -0.008 ~ -0.003 ~ 0.000
In 2 vs Out 1:  0.014 ~ 0.011 ~ 0.002 ~ 0.005 ~ 0.012 ~ 0.021 ~ 0.031
In 3 vs Out 1:  0.000 ~ 0.007 ~ -0.009 ~ -0.004 ~ 0.000 ~ 0.004 ~ 0.010
In 4 vs Out 1:  -0.003 ~ 0.006 ~ -0.011 ~ -0.007 ~ -0.002 ~ 0.001 ~ 0.004
In 5 vs Out 1:  0.011 ~ 0.008 ~ 0.002 ~ 0.004 ~ 0.010 ~ 0.016 ~ 0.023
In 6 vs Out 1:  -0.003 ~ 0.009 ~ -0.015 ~ -0.008 ~ -0.002 ~ 0.002 ~ 0.007
In 7 vs Out 1:  -0.001 ~ 0.009 ~ -0.013 ~ -0.006 ~ 0.000 ~ 0.004 ~ 0.011
In 8 vs Out 1:  -0.007 ~ 0.011 ~ -0.022 ~ -0.014 ~ -0.005 ~ 0.000 ~ 0.005
In 9 vs Out 1:  0.011 ~ 0.011 ~ 0.000 ~ 0.003 ~ 0.009 ~ 0.019 ~ 0.027
In 10 vs Out 1:  -0.005 ~ 0.011 ~ -0.019 ~ -0.011 ~ -0.003 ~ 0.001 ~ 0.008
In 11 vs Out 1:  0.009 ~ 0.005 ~ 0.002 ~ 0.005 ~ 0.008 ~ 0.013 ~ 0.017
```

Figure E-8 Check s Input Sensitivity for each input-output pair